

# Social-Spider Optimization-based Support Vector Machines applied for energy theft detection<sup>☆</sup>



Danillo R. Pereira<sup>a</sup>, Mario A. Pazoti<sup>a</sup>, Luís A.M. Pereira<sup>b</sup>, Douglas Rodrigues<sup>c</sup>,  
Caio O. Ramos<sup>d</sup>, André N. Souza<sup>d</sup>, João P. Papa<sup>e,\*</sup>

<sup>a</sup> Informatics Faculty of Presidente Prudente, The University of Western São Paulo, Presidente Prudente, SP, Brazil

<sup>b</sup> Institute of Computing, University of Campinas, Campinas, SP, Brazil

<sup>c</sup> Department of Computer Science, Federal University of São Carlos, São Carlos, SP, Brazil

<sup>d</sup> Department of Electrical Engineering, São Paulo State University, Bauru, São Paulo, Brazil

<sup>e</sup> Department of Computing, São Paulo State University, Bauru, São Paulo, Brazil

## ARTICLE INFO

### Article history:

Received 14 November 2014

Revised 2 November 2015

Accepted 2 November 2015

Available online 1 December 2015

### Keywords:

Nontechnical losses

Power distribution systems

Social-Spider Optimization

Support Vector Machines

## ABSTRACT

The problem of Support Vector Machines (SVM) tuning parameters (i.e., model selection) has been paramount in the last years, mainly because of the high computational burden for SVM training step. In this paper, we address this problem by introducing a recently developed evolutionary-based algorithm called Social-Spider Optimization (SSO), as well as we introduce SSO for feature selection purposes. The model selection task has been handled in three distinct scenarios: (i) feature selection, (ii) tuning parameters and (iii) feature selection+tuning parameters. Such extensive set of experiments against with some state-of-the-art evolutionary optimization techniques (i.e., Particle Swarm Optimization and Novel Global-best Harmony Search) demonstrated SSO is a suitable approach for SVM model selection, since it obtained the top results in 8 out of 10 datasets employed in this work (considering all three scenarios). Notice the best scenario seemed to be the combination of both feature selection and SVM tuning parameters. In addition, we validated the proposed approach in the context of theft detection in power distribution systems.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Machine learning techniques have been actively pursued in the last decades, since to recognize patterns in different applications through a learning process is of great interest. Based on the statistical learning theory, Support Vector Machines (SVM) [1] are based on the maximal margin assumption, which considers a hyperplane that separates the dataset samples in a high dimensional feature space induced by a non-linear mapping using kernel functions. Although SVM have been considered one of the state-of-the-art pattern recognition techniques, they suffer from the high computational burden for training patterns. Some kernel functions are parameterized, which means there is a need for optimization techniques to find out a suitable set of values that are less prone for classification errors over a training/validating set.

<sup>☆</sup> Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. M. Malek

\* Corresponding author. Tel./fax: +55 14 31036079.

E-mail addresses: [danilopereira@unoeste.br](mailto:danilopereira@unoeste.br) (D.R. Pereira), [mario@unoeste.br](mailto:mario@unoeste.br) (M.A. Pazoti), [luismartinspr@gmail.com](mailto:luismartinspr@gmail.com) (L.A.M. Pereira), [douglasrodrigues.dr@gmail.com](mailto:douglasrodrigues.dr@gmail.com) (D. Rodrigues), [caioramos@gmail.com](mailto:caioramos@gmail.com) (C.O. Ramos), [andrejau@feb.unesp.br](mailto:andrejau@feb.unesp.br) (A.N. Souza), [papa@fc.unesp.br](mailto:papa@fc.unesp.br), [papa.joaopaulo@gmail.com](mailto:papa.joaopaulo@gmail.com) (J.P. Papa).

Although the reader can face several works that deal with new approaches for SVM modeling and parameter optimization, in this work we focus on evolutionary strategies for the latter purpose. Friedrichs and Igel [2], for instance, presented an evolutionary-based approach for SVM parameter optimization using the Covariance Matrix Adaptation Evolution Strategy. Howley and Madden [3] proposed a kernel optimization method based on Genetic Programming, and Lessmann et al. [4] employed the well-known Genetic Algorithm for the same task. In regard to Particle Swarm Optimization (PSO)-based SVM training, one can be referred to several works: Liu et al. [5], for instance, proposed an integrated approach that aimed to optimize both features and the parameters of a Support Vector Machines classifier. Melgani and Bazi [6] presented an SVM model selection approach based on PSO for Electrocardiogram signal classification; in this work, only SVM parameters have been optimized. More recently, Pereira et al. [7] introduced the Harmony Search algorithm for training SVM classifiers in the context of theft detection in power distribution systems, and Cawley [8] proposed an approach based on Tabu Search for model selection in SVM classifiers.

Based on the social dynamics of spiders, Cuevas et al. [9] proposed the Social-Spider Optimization (SSO), which considers both male and female spiders as well as their cooperative behavior for solving optimization tasks. Such technique has demonstrated very promising results, being also as efficient as some state-of-the-art evolutionary-based approaches. The main contributions of this paper are two-fold: (i) to extend the work by Pereira et al. [10], which introduced SSO for SVM parameter estimation, and (ii) to apply feature selection by means of SSO together with SVM model selection. The proposed SSO-SVM technique is evaluated in public datasets, as well as in the context of theft detection in power distribution systems for the first time. For this latter purpose, we have employed two private datasets provided by a Brazilian electrical power company, which contain legal and illegal profiles from commercial and industrial consumers.

The remainder of this paper is organized as follows. Sections 2 and 3 present the theory background regarding SSO and the methodology employed in this work, respectively. Experiments are described in Section 4, and the conclusions are stated in Section 5.

## 2. Social-Spider Optimization

Social-Spider Optimization is based on the cooperative behavior of social-spiders and it was proposed by Cuevas [9]. The algorithm takes into account two genders of search spiders: males and females. Depending on the gender, each agent is conducted by a set of different operators emulating a cooperative behavior in a colony. The search space is assumed as a communal web and a spider's position represents an optimal solution.

An interesting characteristic of social-spiders is the female-biased population. The number of male spiders hardly reaches 30% of the total colony members. The number of females  $N_f$  is randomly selected within a range of 65–90% of the entire population  $N$ , being calculated as follows:

$$N_f = \lfloor (0.9 - rand * 0.25) * N \rfloor, \quad (1)$$

where  $rand$  is a random number between  $[0, 1]$ , thus guaranteeing the aforementioned range considering the number of female spiders. The number of male spiders  $N_m$  is given by:

$$N_m = N - N_f. \quad (2)$$

Every spider receives a weight according to the fitness value of the solution:

$$w_i = \frac{fitness_i - worst}{best - worst}, \quad (3)$$

where  $fitness_i$  is the fitness value obtained by the evaluation of the  $i$ th spider's position  $i = 1, 2, \dots, N$ . The *worst* and *best* mean the worst fitness value and best fitness value of the entire population, respectively.

The communal web is used as a mechanism to transmit information among the colony members. The information is encoded as small vibrations and depends on the weight and distance of the spider which has generated them:

$$V_{i,j} = w_j * e^{-d_{i,j}^2}, \quad (4)$$

where  $d_{i,j}$  is the Euclidean Distance between the spider  $i$  and  $j$ . We can consider three special relationships:

- Vibrations  $V_{i,c}$  are perceived by the spider  $i$  as a result of the information transmitted by the member  $c$  who is the nearest member to  $i$ , and possesses a higher weight  $w_c > w_i$ ;
- The vibrations  $V_{i,b}$  perceived by the spider  $i$  as a result of information transmitted by the spider  $b$  holding the best weight of the entire population;
- The vibrations  $V_{i,f}$  perceived by the spider  $i$  as a result of the information transmitted by the nearest female  $f$ .

Social-spiders perform cooperative interaction over other colony members depending on the gender. In order to emulate the cooperative behavior of the female spider, a new operator is defined in Eq. (5). The movement of attraction or repulsion is

developed over other spiders according to their vibrations, which are emitted over the communal web:

$$f_i = \begin{cases} f_i + \alpha * V_{i,c} * (s_c - f_i) + \beta * V_{i,b} * \\ (s_b - f_i) + \gamma * \left( rand - \frac{1}{2} \right) \\ \text{if } r_m < PF; \\ f_i - \alpha * V_{i,c} * (s_c - f_i) - \beta * V_{i,b} * \\ (s_b - f_i) + \gamma * \left( rand - \frac{1}{2} \right) \\ \text{if } r_m \geq PF, \end{cases} \quad (5)$$

where  $r_m$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $rand$  are uniform random numbers between  $[0, 1]$ , and  $s_c$  and  $s_b$  represent the nearest member to  $i$  that holds a higher weight and the best spider of the entire population, respectively. The input variable  $PF \in [0, 1]$  is thus used to control whether the current spider's position will be updated with a positive or negative direction.

The male spider population is divided into two classes: dominant and non-dominant. The dominant class spider has better fitness in comparison to non-dominant, and they are attracted to the closest female spider in the communal web. In the other hand, non-dominant male spiders tend to concentrate in the center of the male population as a strategy to take advantage of resources that are wasted by dominant males:

$$m_i = \begin{cases} m_i + \alpha * V_{i,f} * (s_f - m_i) + \gamma * \left( rand - \frac{1}{2} \right) \\ \text{if } w_{N_f+i} > w_{N_f+m}; \\ m_i + \alpha * \left( \frac{\sum_{h=1}^{N_m} m_h * w_{N_f+h}}{\sum_{h=1}^{N_m} w_{N_f+h}} \right) \\ \text{if } w_{N_f+i} \geq w_{N_f+m}. \end{cases} \quad (6)$$

where  $s_f$  represents the nearest female spider to the male spider  $i$ .

Mating is performed by dominant males and female members in a social-spider colony. Considering  $r$  (calculated by Eq. 7) as being the radius, when a dominant male spider locates female members inside  $r$ , it mates, forming a new brood:

$$r = \frac{\sum_{j=1}^n p_j^{high} - p_j^{low}}{2 * n}, \quad (7)$$

where  $n$  is the dimension of the problem.  $p_j^{high}$  and  $p_j^{low}$  are the upper and lower bounds, respectively. Once the new spider is formed, it is compared to the worst spider of the colony. If the new spider is better, the worst spider is replaced by the new one.

### 3. Methodology

In this section, we present the methodology used to assess the robustness and efficiency of SSO against with two others commonly used optimization methods: PSO and NGHS (Novel Global Harmony Search) [11]. Such approaches have been selected for comparison purposes since they have been already applied in the same context of this paper, as stated in Section 1.

The experimental setup has been divided in three different stages: (i) feature selection; (ii) parameters tuning; and (iii) parameters tuning combined with feature selection: this experiment has been conducted applying the SVM parameters tuning first, for further feature selection. Thus, the tuning step is performed prior to the feature selection one, which employs the optimum set of SVM parameters already computed. In addition, for each aforementioned experiment, we have executed the optimization approaches using two different SVM kernels: RBF and Polynomial.

In order to validate the experiments, we employed 10 public benchmarking datasets<sup>1</sup> that have been frequently used for the evaluation of different optimization methods. Table 1 presents the main characteristics of each dataset, which has been randomly partitioned in three sets: training ( $Z_1$ ), evaluating ( $Z_2$ ) and testing ( $Z_3$ ) sets, with 40%, 30% and 30% of the total samples, respectively (such values have been empirically set). The datasets have been chosen in order to represent distinct scenarios, which comprise datasets with different number of features, sizes and classes.

In this work, we adopted the following procedure: SVM is trained over  $Z_1$  and its accuracy over  $Z_2$  is then used as the fitness value for each particle (spider, particle or harmonies). After the convergence process of the optimization algorithm, the final SVM classifier is employed over the unseen set  $Z_3$ . Such procedure is conducted using a cross-validation with 10 runnings. In order to compare the optimization methods, we computed the mean accuracy and execution times (seconds) for each of them. In addition, we employed the Wilcoxon test [12] to rank the techniques, as well as to provide a more robust statistical evaluation.

**Table 1**

Information about the benchmarking datasets used in the experiments.

Dataset	#samples	#features	#classes
glass	214	9	6
fourclass	862	2	2
heart	270	13	2
iris	150	4	3
liver_disorders	345	6	2
australia	690	14	2
ionosphere	351	34	2
sonar	208	60	2
vehicle	846	18	4
diabetes	768	8	2

**Table 2**

The parameters used by each optimization algorithm.

Algorithm	Parameters
PSO	$c_1 = c_2 = 2.0$ and $w = 0.9$
SSO	$PF = 0.5$
NGHS	$p_m = 0.1$

**Table 3**

Fixed parameters used for “Standard” configuration.

Kernel	Parameters
RBF	$\gamma = 0.25$ and $C = 1.2$
Polynomial	$C = 0.9$ , $\gamma = 0.8$ , $d = 5$ and $c = 0.9$

The parameters used by each meta-heuristic optimization algorithm are displayed in Table 2. It is important to shed light over that such values have been empirically set and based on previous experiments [7,9,13], as well as such values are quite similar to the ones recommended by their proponents.

In regard to SVM source code, we used the well-known LibSVM [14] package, and for PSO and HS approaches we used our own implementation. In order to allow a fair comparison among all techniques, we have provided the results without feature selection and parameters tuning, denoted here as “Standard”. For this experiment, only the training ( $Z_1$ ) and test ( $Z_3$ ) sets have been used. The fixed parameters used for different kernels to obtain the “Standard” configuration are presented in Table 3, being such values empirically chosen. Finally, we evaluated SSO in the context of theft detection in power distribution systems in the three aforementioned stages of the experimental setup.

## 4. Experiments

In this section, we present the results obtained for the three different stages of the experimental setup.

### 4.1. Feature selection

This section presents the results obtained by PSO, NGHS and SSO for feature selection purposes. The following sections state the experimental results applied in two different SVM kernels: RBF and Polynomial. Notice for all optimization algorithms we used 20 agents and 200 iterations (such values have been empirically set).

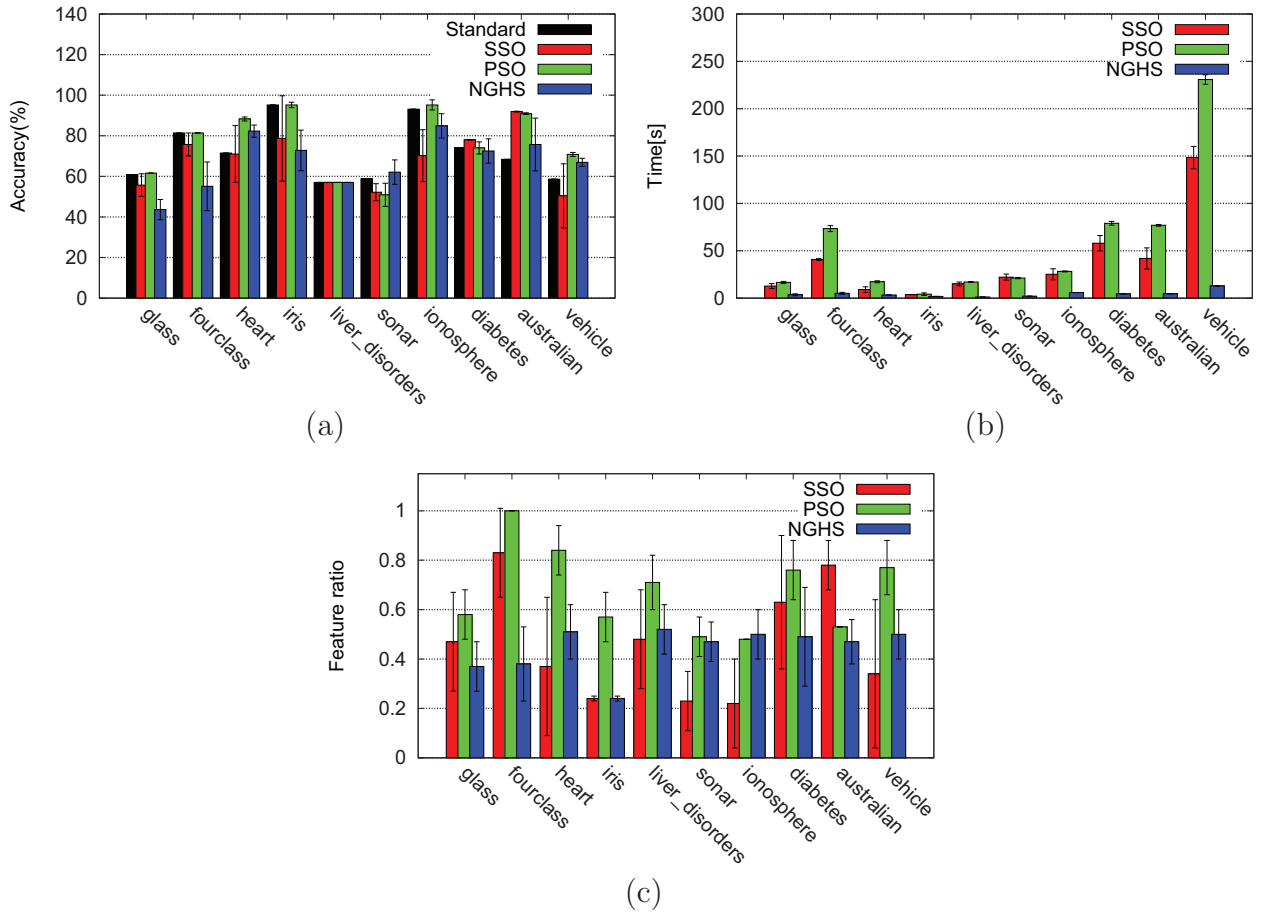
#### 4.1.1. RBF

Fig. 1 a and b shows the mean accuracy and execution times, respectively, considering the RBF kernel. Table 4 presents the Wilcoxon statistical evaluation for the feature selection experiment considering the RBF kernel.

From the accuracy and execution time experiments, we can conclude PSO has been the most accurate technique, followed by SSO and NGHS. However, NGSH has been the fastest technique, followed by SSO and PSO. Fig. 1c displays the feature selection ratio for all techniques, in which one can observe PSO has allowed a better reduction ratio, except for “australian” dataset (SSO has provided the best reduction ratio).

The lower recognition rates achieved by NGHS might be the result of a small number of iterations, which may halt the algorithm during its convergence process. Except for “sonar” dataset, NGHS did not perform well. Although it has been the fastest technique, since it updates one agent at each iteration only, approaches based on Harmony Search can be affected by a small

<sup>1</sup> <http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/>.



**Fig. 1.** Feature selection experiment using RBF kernel: (a) mean accuracy and standard deviation considering “Standard” configuration, SSO, PSO and NGHS. (b) Mean execution times [s] obtained by SSO, PSO and HS. (c) Feature selection ratio through SSO, PSO and NGHS.

**Table 4**

Wilcoxon statistical analysis for RBF kernel: symbol “≠” denotes there exists difference between methods, and the symbol “=” states the techniques are similar each other.

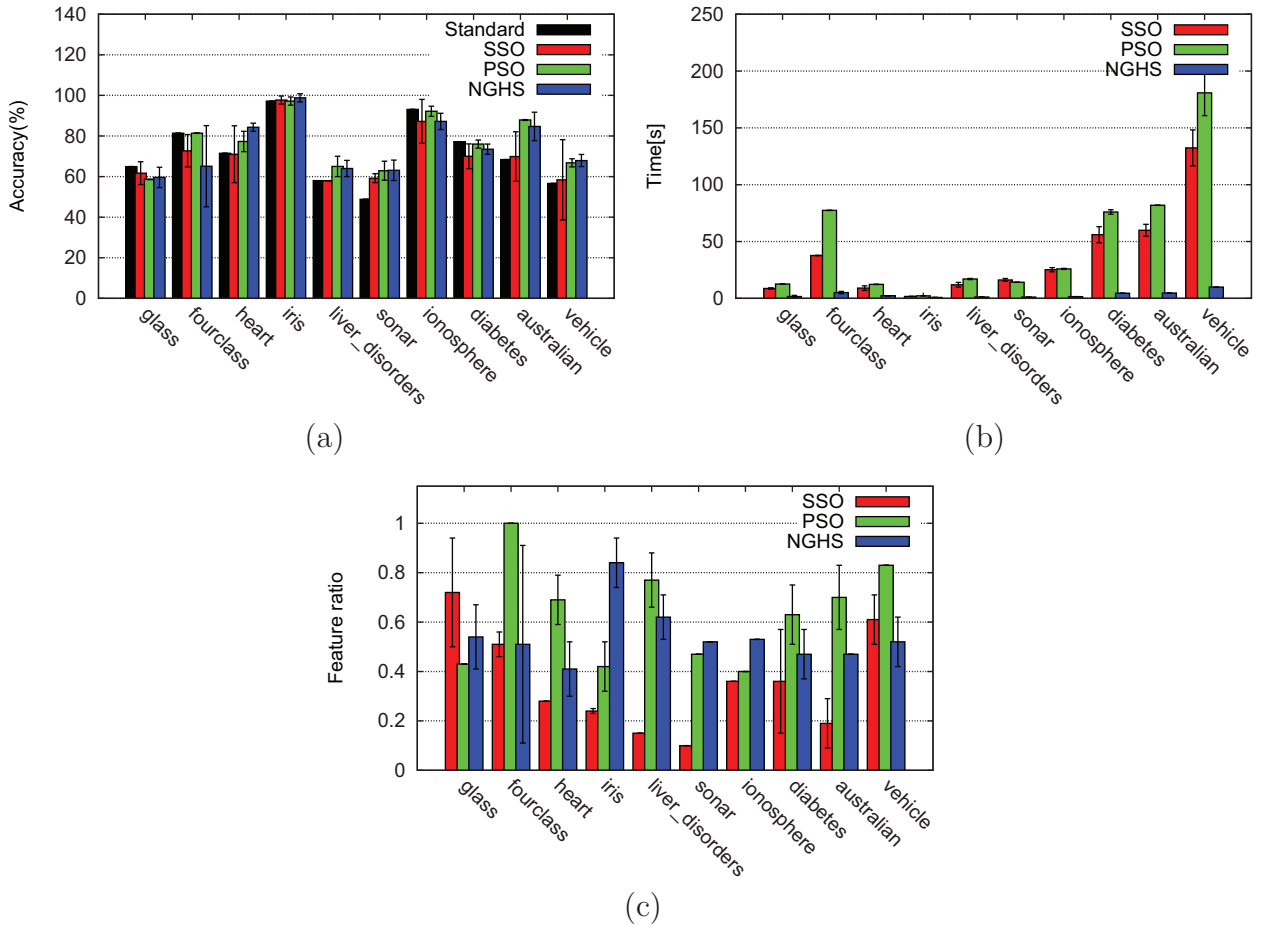
Dataset	Wilcoxon test	
	SSO/PSO	SSO/NGHS
glass	≠	≠
fourclass	≠	≠
heart	≠	≠
iris	≠	≠
liver_disorders	=	=
sonar	=	≠
ionosphere	≠	≠
diabetes	=	≠
australian	=	≠
vehicle	≠	≠

number of iterations. Indeed, it is not straightforward to establish a fair number of iterations for all meta-heuristics that consider their shortcomings individually.

4.1.2. Polynomial kernel

Fig. 2 a and b shows the mean accuracy and execution times, respectively, considering the Polynomial kernel. Table 5 presents the Wilcoxon statistical evaluation for the feature selection experiments in regard to the Polynomial kernel.

From the accuracy and execution times experiments, we can observe SSO has outperformed PSO and NGHS for “glass” dataset, and for the remaining datasets the best techniques have been PSO and NGHS. Once again, NGHS has been the fastest technique,



**Fig. 2.** Feature selection experiment using Polynomial kernel: (a) mean accuracy and standard deviation considering “Standard” configuration, SSO, PSO and NGHS. (b) Mean execution times [s] obtained by SSO, PSO and HS. (c) Feature selection ratio through SSO, PSO and NGHS.

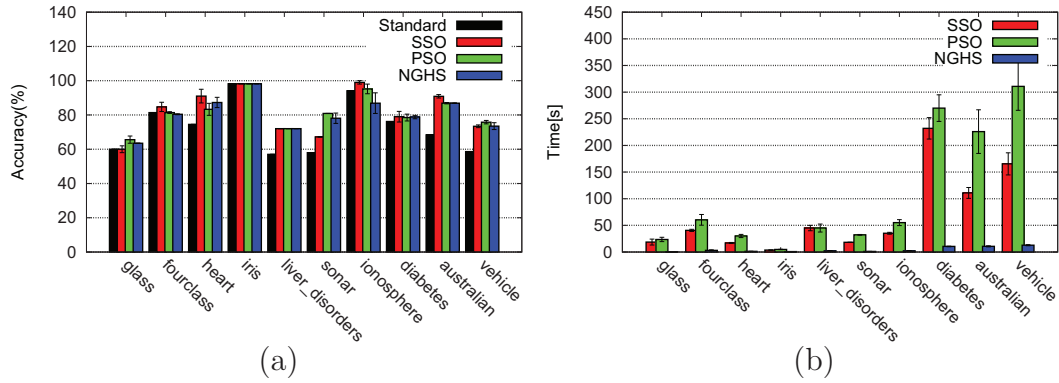
**Table 5**

Wilcoxon statistical analysis for Polynomial kernel: symbol “≠” denotes there exists difference between methods, and the symbol “=” states the techniques are similar each other.

Dataset	Wilcoxon test	
	SSO/PSO	SSO/NGHS
glass	≠	=
fourclass	≠	≠
heart	≠	≠
iris	=	≠
liver_disorders	≠	≠
sonar	=	≠
ionosphere	≠	=
diabetes	≠	≠
australian	≠	≠
vehicle	≠	≠

followed by SSO. Fig. 2c displays the feature selection ratio for all techniques, in which one can observe SSO achieved the best reduction rate for “glass” dataset, being PSO the one with the best reduction rates for “fourclass”, “heart”, “liverdisorders”, “diabetes”, “australian” and “vehicle” datasets.

In regard to the feature selection experiment, PSO has been the most accurate technique for SVM with RBF kernel (for a considerable number of datasets), followed by SSO and NGHS. If we consider the computational load, NGHS is the fastest technique, followed by SSO and PSO. Notice this same behavior can be found for SVM with Polynomial kernel, with NGHS achieving best results than the ones obtained through RBF kernel.



**Fig. 3.** SVM parameters tuning experiment using RBF kernel: (a) recognition rates considering “Standard” configuration, SSO, PSO and NGHS. (b) Computational load [s] for SSO, PSO and NGHS techniques.

**Table 6**

Wilcoxon statistical analysis for RBF kernel: symbol “ $\neq$ ” denotes there exists difference between the methods, and the symbol “=” represents the techniques are similar each other.

Dataset	Wilcoxon test	
	SSO/PSO	SSO/NGHS
glass	$\neq$	$\neq$
fourclass	=	$\neq$
heart	$\neq$	$\neq$
iris	=	=
liver_disorders	=	=
sonar	$\neq$	$\neq$
ionosphere	$\neq$	$\neq$
diabetes	=	=
australian	$\neq$	=
vehicle	$\neq$	=

## 4.2. Parameters tuning

This section presents the results obtained by PSO, NGHS and SSO for SVM tuning parameters. The following sections state the experimental results applied in two different SVM kernels: RBF and Polynomial. Notice for all optimization algorithms we used 20 agents and 200 iterations (such values have been empirically set).

### 4.2.1. RBF kernel

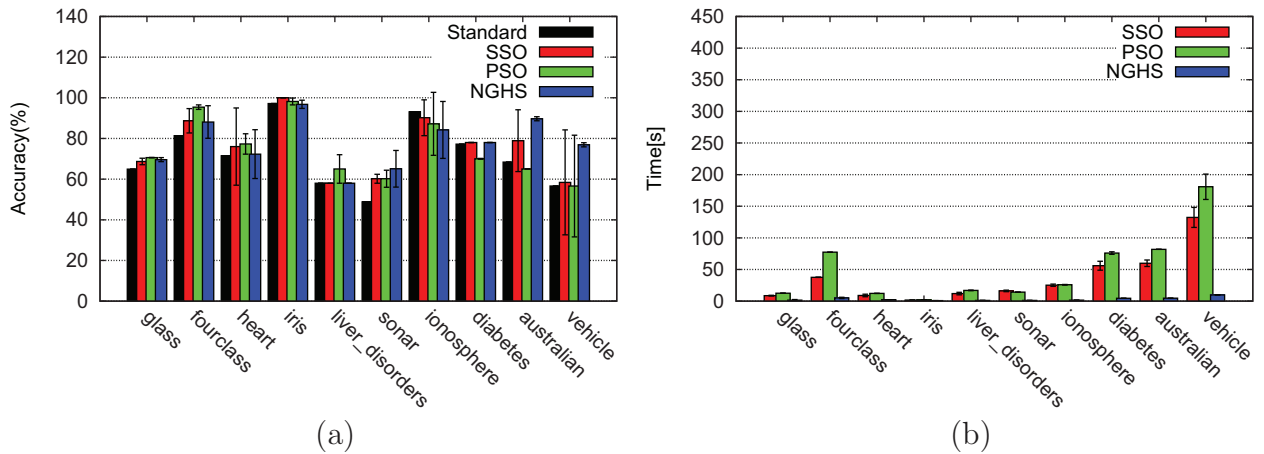
The searching range of parameter  $C$  was defined within the interval  $[-32, 32]$ , while the searching range of parameter  $\gamma$  was within the interval  $[0, 32]$ . These ranges were chosen based on our previous experience. Fig. 3a and b displays the mean accuracy and the mean execution time (seconds) for each optimization technique, respectively. Table 6 presents the Wilcoxon statistical evaluation for the SVM tuning parameters experiment considering RBF kernel.

If we consider the Wilcoxon statistical evaluation, it is possible to observe SSO has outperformed PSO and NGHS for “heart” and “ionosphere” datasets, and obtained similar results to PSO for “fourclass” dataset. In regard to “iris” and “live\_disorders” datasets, all techniques have been similar each other. Therefore, we can observe SSO has obtained the top results in 7 out of 10 datasets for SVM tuning parameters with RBF kernel. Once again, the computational load behavior has been the same as before, i.e., NGHS is the fastest approach, followed by SSO and PSO.

### 4.2.2. Polynomial kernel

The searching range of parameters was defined within the intervals:  $C \in [0, 2]$ ,  $\gamma \in [0, 2]$ ,  $c \in [0, 10]$  and  $d \in [0, 2]$ . These ranges were chosen based on our previous experience. Fig. 4a and b displays the mean accuracy and the mean execution time (seconds) for each optimization technique, respectively. Table 7 presents the Wilcoxon statistical evaluation for the SVM tuning parameters experiment considering the Polynomial kernel.

Considering the Wilcoxon test, SSO has outperformed PSO and NGHS for “iris” dataset. In regard to “ionosphere” and “diabetes” datasets, SSO has outperformed PSO and NGHS techniques, respectively. Considering the execution time, NGHS has been the fastest approach, followed by SSO and PSO.



**Fig. 4.** SVM parameters tuning experiment using Polynomial kernel: (a) recognition rates considering “Standard” configuration, SSO, PSO and NGHS. (b) Computational load [s] for SSO, PSO and NGHS techniques.

**Table 7**

Wilcoxon statistical analysis for Polynomial kernel: symbol “≠” denotes there exists difference between the methods, and the symbol “=” represents the techniques are similar each other.

Dataset	Comparison	
	SSO/PSO	SSO/NGHS
glass	≠	=
fourclass	≠	≠
heart	≠	≠
iris	≠	≠
liver_disorders	≠	=
sonar	=	≠
ionosphere	=	≠
diabetes	≠	=
australian	≠	≠
vehicle	≠	≠

In regard to the SVM tuning parameters experiment, SSO has obtained very interesting results considering RBF kernel: it has outperformed PSO and NGHS in 70% of the employed datasets considering the aforementioned kernel. If we consider the computational load, NGHS has been the fastest approach, followed by SSO and PSO.

#### 4.3. Parameters tuning combined with feature selection

This section presents the results obtained by PSO, NGHS and SSO for SVM tuning parameters combined with feature selection. The following sections state the experimental results applied in two different SVM kernels: RBF and Polynomial. Notice for all optimization algorithms we used 10 agents and 100 iterations (such values have been empirically set). It is important to highlight to the reader we have tried to use 20 agents and 200 iterations as before. However, the difference in accuracy did not justify the computational load for that.

##### 4.3.1. RBF kernel

Fig. 5 a and b shows the mean accuracy and execution times, respectively, considering the RBF kernel. Table 8 presents the Wilcoxon statistical evaluation for the SVM tuning parameters combined with feature selection experiments in regard to the RBF kernel.

If we consider the Wilcoxon statistical analysis, SSO has outperformed PSO and NGHS for “liver\_disorders”, “australian” and “vehicle” datasets. In addition, SSO has achieved better results than NGHS for “glass”, “fourclass”, “heart” and “ionosphere” datasets, being such results similar to the ones obtained by PSO in these datasets. Therefore, SSO has obtained the top results in 7 out of 10 datasets. In regard to the computational load, NGHS has been the fastest technique, followed by SSO and PSO. It is interesting to stress that SSO feature selection ratios (Fig. 5c) have been much better than the ones obtained in the “single feature selection” experiment (Fig. 1c), which evidences the benefits of a combined optimization between feature selection and parameters tuning.



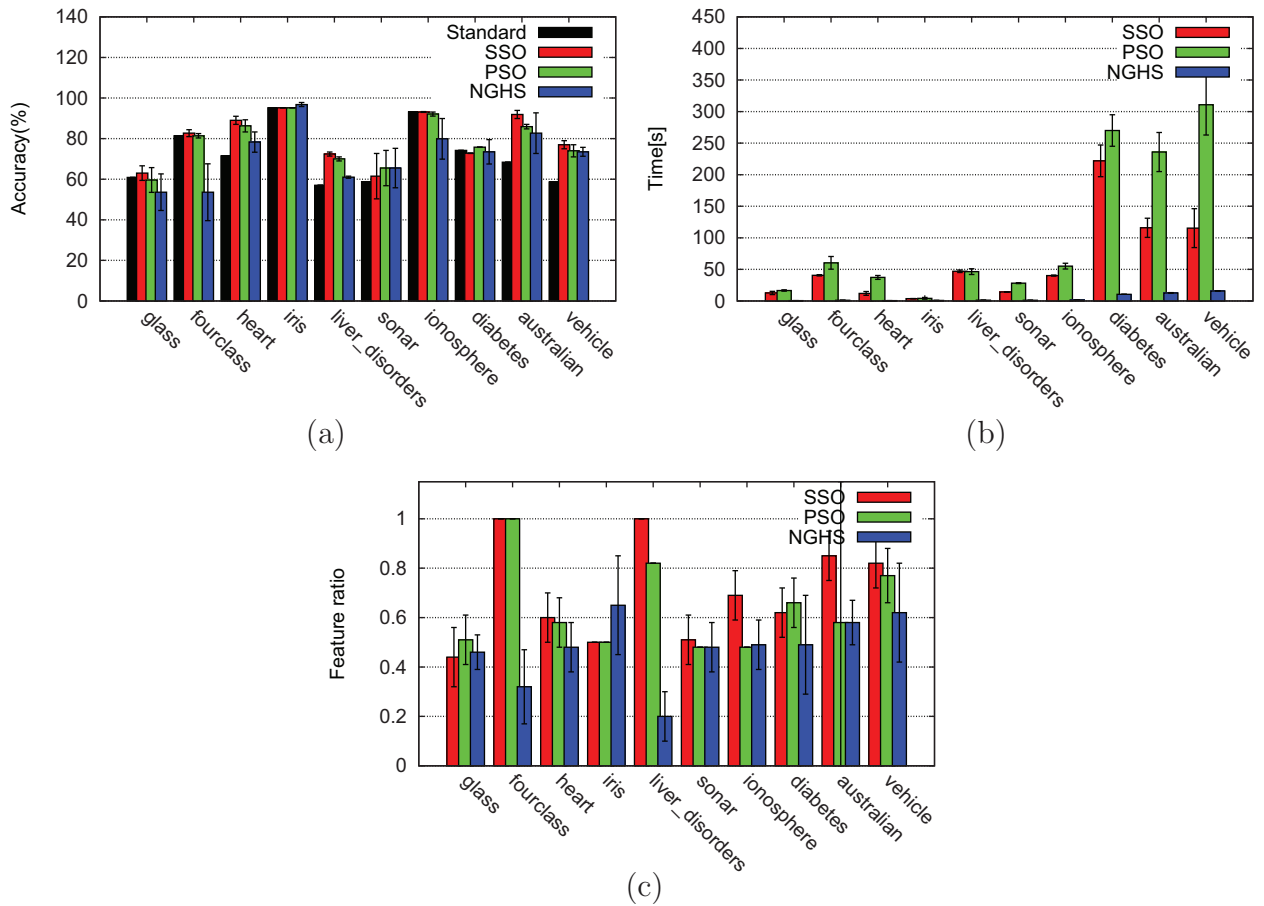


Fig. 5. Feature selection combined with SVM parameters tuning experiment using RBF kernel: (a) mean accuracy and standard deviation considering "Standard" configuration, SSO, PSO and NGHS. (b) Mean execution times [s] obtained by SSO, PSO and HS. (c) Feature selection ratio through SSO, PSO and NGHS.

Table 8

Wilcoxon statistical analysis for RBF kernel: symbol " ≠ " denotes there exists difference between the methods, and the symbol "=" represents the techniques are similar each other.

Dataset	Wilcoxon test	
	SSO/PSO	SSO/NGHS
glass	=	≠
fourclass	=	≠
heart	=	≠
iris	=	=
liver_disorders	≠	≠
sonar	≠	≠
ionosphere	=	≠
diabetes	≠	≠
australian	≠	≠
vehicle	≠	≠

#### 4.3.2. Polynomial kernel

Fig. 6 a and b shows the mean accuracy and execution times, respectively, considering the Polynomial kernel. Table 9 presents the Wilcoxon statistical evaluation for the SVM tuning parameters combined with feature selection experiments in regard to the Polynomial kernel.

The experimental results combining SVM parameters tuning and feature selection does not seem to improve the original results obtained using only feature selection (Fig. 2) and only SVM parameter tuning (Fig. 4) considering the Polynomial kernel. It seems Polynomial kernel cannot improve the results over these datasets, since its results are inferior to the ones obtained by RBF kernel, which has been the most accurate one.

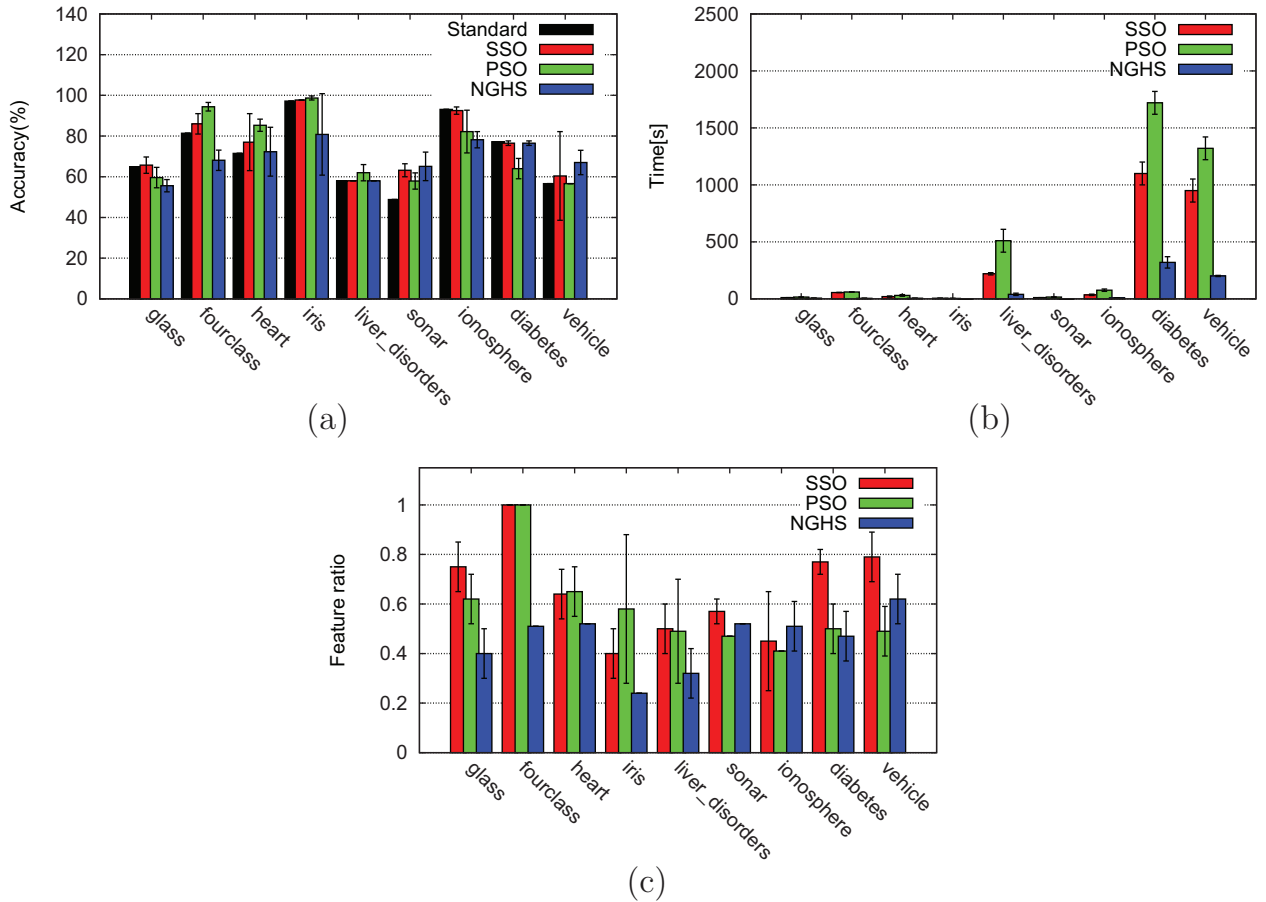


Fig. 6. Feature selection combined with SVM parameters tuning experiment using the Polynomial kernel: (a) mean accuracy and standard deviation considering “Standard” configuration, SSO, PSO and NGHS. (b) Mean execution times [s] obtained by SSO, PSO and HS. (c) Feature selection ratio through SSO, PSO and NGHS.

Table 9

Wilcoxon statistical analysis for Polynomial kernel: symbol “≠” denotes there exists difference between the methods, and the symbol “=” represents the techniques are similar each other.

Dataset	Comparison	
	SSO/PSO	SSO/NGHS
glass	≠	≠
fourclass	≠	≠
heart	≠	=
iris	≠	≠
liver_disorders	≠	=
sonar	=	≠
ionosphere	≠	≠
diabetes	≠	=
vehicle	≠	≠

The experiments using the Polynomial kernel did not seem to improve a lot the results for all three experiments, being the results using only SVM parameter tuning the best ones for almost all cases.

4.4. Discussion

In this section, we presented a deeper analysis regarding all aforementioned experiments. Tables 10 and 11 show the mean accuracy obtained by “feature selection”, “parameters tuning” and “parameters tuning and feature selection” using RBF kernel. Notice the values in bold mean the best ones considering the kernel itself, and the “underlined” values marked stand for the best ones considering all kernels. It is possible to observe SSO has obtained the best results (sometimes similar) in 8 out of 10 datasets

**Table 10**

Mean classification rates using feature selection and SVM parameters tuning considering RBF kernel (the best values are in bold).

Dataset	Feature Selection Accuracy				Parameters Tuning Accuracy		
	Standard	SSO	PSO	NGHS	SSO	PSO	NGHS
glass	60.46	56.5 ± 8.7	62.79 ± 0.0	41.86 ± 6.8	60.46 ± 1.3	<b>65.89 ± 1.0</b>	65.11 ± 0.0
fourclass	80.81	75.38 ± 7.6	80.81 ± 0.0	54.84 ± 13.7	<b>83.72 ± 1.9</b>	82.36 ± 1.1	81.97 ± 0.8
heart	74.07	72.83 ± 14.3	<b>90.74 ± 1.4</b>	83.33 ± 4.0	<b>90.74 ± 2.6</b>	84.56 ± 2.3	87.03 ± 2.6
iris	96.66	78.88 ± 25.1	96.66 ± 1.4	73.33 ± 22.2	96.66 ± 0.0	96.66 ± 0.0	96.66 ± 0.0
liver_disorders	57.97	57.97 ± 0.0	57.97 ± 0.0	57.97 ± 0.0	<b>72.46 ± 0.0</b>	<b>72.46 ± 0.0</b>	<b>72.46 ± 0.0</b>
sonar	58.06	52.68 ± 5.4	51.61 ± 6.9	61.29 ± 4.5	67.74 ± 0.0	<b>80.64 ± 0.0</b>	78.49 ± 3.0
ionosphere	95.08	72.13 ± 11.5	96.17 ± 2.0	85.24 ± 7.0	<b>98.36 ± 1.6</b>	96.17 ± 1.5	87.43 ± 7.6
diabetes	75.75	78.78 ± 0.0	77.27 ± 1.0	74.49 ± 4.9	<b>79.54 ± 2.0</b>	79.04 ± 1.2	79.29 ± 0.7
australian	68.90	89.07 ± 0.6	88.79 ± 0.7	76.19 ± 10.2	89.91 ± 0.9	87.67 ± 0.3	87.95 ± 0.3
vehicle	57.93	50.57 ± 14.6	71.26 ± 0.3	67.58 ± 2.5	73.79 ± 0.3	75.40 ± 0.6	74.25 ± 1.4

**Table 11**

Mean classification rates for SVM parameters tuning combined with feature selection considering RBF kernel (the best values are in bold).

Dataset	Parameters tuning and feature selection Accuracy		
	SSO	PSO	NGHS
glass	62.79 ± 3.0	59.68 ± 5.4	53.48 ± 9.8
fourclass	<b>83.72 ± 1.1</b>	82.36 ± 1.1	54.84 ± 13.7
heart	88.88 ± 1.9	86.41 ± 2.3	78.39 ± 5.7
iris	96.66 ± 0.0	96.66 ± 1.4	<b>97.77 ± 1.5</b>
liver_disorders	<b>72.46 ± 1.3</b>	70.04 ± 0.6	60.86 ± 0.0
sonar	61.29 ± 11.1	65.59 ± 7.6	65.59 ± 9.2
ionosphere	95.08 ± 0.2	94.53 ± 0.7	79.23 ± 10.0
diabetes	74.24 ± 0.0	77.77 ± 0.3	74.24 ± 6.5
australian	<b>91.59 ± 2.0</b>	85.15 ± 1.0	82.07 ± 10.5
vehicle	<b>77.24 ± 2.0</b>	74.02 ± 3.2	73.33 ± 2.5

**Table 12**

Mean classification rates using feature selection and SVM parameters tuning considering the Polynomial kernel (the best values are in bold).

Dataset	Feature selection Accuracy				Parameters tuning Accuracy		
	Standard	SSO	PSO	NGHS	SSO	PSO	NGHS
glass	65.11	64.34 ± 4.3	58.13 ± 0.0	59.68 ± 5.8	68.21 ± 1.0	<b>69.76 ± 0.0</b>	68.99 ± 1.0
fourclass	81.39	73.83 ± 6.5	81.39 ± 0.0	65.11 ± 21.0	87.98 ± 4.4	<b>95.54 ± 1.5</b>	87.40 ± 7.7
heart	72.22	72.22 ± 13.6	77.77 ± 5.2	85.18 ± 1.5	76.54 ± 14.8	83.33 ± 5.2	72.83 ± 13.0
iris	96.66	96.66 ± 1.4	96.66 ± 1.4	98.88 ± 1.5	<b>100.0 ± 0.0</b>	98.88 ± 1.5	97.77 ± 1.5
liver_disorders	57.97	57.97 ± 0.0	<b>65.70 ± 5.4</b>	63.28 ± 3.4	57.97 ± 0.0	<b>65.70 ± 5.4</b>	57.97 ± 0.0
sonar	48.38	59.13 ± 1.5	61.29 ± 4.5	62.36 ± 4.0	60.21 ± 1.5	60.21 ± 5.4	64.51 ± 9.4
ionosphere	<b>95.08</b>	85.79 ± 11.9	94.53 ± 2.7	86.33 ± 3.3	90.16 ± 9.2	86.33 ± 15.8	84.15 ± 14.3
diabetes	77.27	69.19 ± 5.7	76.51 ± 1.8	73.73 ± 4.4	<b>77.77 ± 0.3</b>	69.84 ± 0.0	77.52 ± 0.3
australian	68.90	70.02 ± 14.0	86.55 ± 0.0	83.75 ± 8.1	77.87 ± 15.8	64.53 ± 3.3	<b>89.07 ± 1.1</b>
vehicle	55.86	58.39 ± 19.3	74.94 ± 2.3	67.35 ± 3.4	58.39 ± 22.8	55.51 ± 0.0	<b>75.86 ± 0.9</b>

(considering all sort of experiments, i.e., “feature selection”, “parameters tuning” and “parameters tuning and feature selection”). The results obtained with “parameters tuning” have been the best among the other two experiments, followed by “parameters tuning and feature selection” and “feature selection” itself. It seems the feature selection combined with SVM parameters tuning have enhanced the results of purely applying feature selection.

Tables 12 and 13 show the mean accuracy obtained by “Feature selection”, “parameters tuning” and “parameters tuning and feature selection” using the Polynomial kernel. In this case, SSH has outperformed PSO and NGHS in 2 out of 10 datasets, being the results using only SVM parameter optimization the best ones.

If we consider the best results among all kernels (“underlined” values), it is possible to observe the RBF kernel has been the most accurate one, achieving the top results in 5 out of 10 datasets. Considering such results, SSO has outperformed PSO and NGHS in 2 out of 5 datasets, being also similar to PSO in one dataset (“liver\_disorders”). A more careful look at the results can lead us to the following statement: SSO has obtained the best results in around 50% of the datasets (“heart/RBF kernel”, “liver\_disorders/RBF kernel”, “ionosphere/RBF kernel”, “diabetes/RBF kernel” and Polynomial kernels”) considering the best global values, i.e., the ones which consider all kernels. Such fact, allied with the SSO efficiency (it has been consistently the second fastest approach in all experiments), makes it a very interesting evolutionary-based approach for SVM parameters tuning.

**Table 13**

Mean classification rates for SVM parameters tuning considering the Polynomial kernel (the best values are in bold). The results marked with \*\*\* mean the LibSVM did not reach a suitable hyperplane with the maximum number of iterations set in this work.

Dataset	Parameters tuning and feature selection		
	Accuracy		
	SSO	PSO	NGHS
glass	65.89 ± 3.9	59.68 ± 6.6	54.26 ± 2.9
fourclass	87.98 ± 4.4	<b>95</b>	69.18 ± 6.5
heart	77.16 ± 15.2	<b>85.80 ± 3.1</b>	73.45 ± 13.2
iris	96.66 ± 0.0	97.77 ± 1.5	79.99 ± 23.5
liver_disorders	57.97 ± 0.0	61.35 ± 4.7	57.97 ± 0.0
sonar	62.36 ± 3.0	56.98 ± 4.0	<b>65.59 ± 9.2</b>
ionosphere	94.53 ± 1.5	81.96 ± 12.7	78.68 ± 13.5
diabetes	76.76 ± 0.9	64.84 ± 5.9	76.76 ± 0.7
australian	***	***	***
vehicle	60.22 ± 23.1	55.51 ± 0.0	65.97 ± 5.0

**Table 14**

Mean classification rates using feature selection for RBF kernel (the best values are in bold).

Dataset	Accuracy		
	SSO	PSO	NGHS
Commercial	97.59 ± 0.32	<b>98.35 ± 0.0</b>	95.38 ± 0.28
Industrial	97.25 ± 1.79	<b>98.82 ± 1.00</b>	95.20 ± 0.87

**Table 15**

Mean classification rates using SVM tuning parameters for RBF kernel (the best values are in bold).

Dataset	Accuracy		
	SSO	PSO	NGHS
Commercial	<b>94.81 ± 0.0</b>	<b>94.81 ± 0.0</b>	<b>94.81 ± 0.0</b>
Industrial	<b>94.50 ± 0.0</b>	<b>94.50 ± 0.0</b>	<b>94.5 ± 0.0</b>

**Table 16**

Mean classification rates using SVM tuning parameters combined with feature selection for RBF kernel (the best values are in bold).

Dataset	Accuracy		
	SSO	PSO	NGHS
Commercial	97.87 ± 0.0	<b>98.58 ± 0.16</b>	96.34 ± 1.67
Industrial	<b>99.81 ± 0.0</b>	99.45 ± 0.0	97.31 ± 1.97

#### 4.5. Theft detection in power distribution systems

In this section, we present the performance of SSO in a real problem of theft detection. The main idea is to identify illegal consumers in power distribution systems using two private datasets from a Brazilian electrical power company containing commercial and industrial profiles. Each profile (industrial and commercial) is represented by eight numerical features, according to the work by Ramos et al. [15]: Demand Billed, Demand Contracted, Demand Measured or Maximum Demand, Reactive Energy, Power Transformer, Power Factor, Installed Power and Load Factor.

In regard to the “feature selection” experiment, we used 20 agents and 200 iterations for all optimization approaches. Table 14 displays the accuracy results. For sake of space, we have discussed here only the results considering RBF kernel, since they have been the most accurate ones. If we consider the standard deviation, it is possible to observe that SSO and PSO have obtained similar recognition rates for “Industrial” dataset.

With respect to the “SVM tuning parameters” experiment, we used 20 agents and 200 iterations for all optimization techniques. Table 15 displays the accuracy results. As the reader can observe, all techniques have achieved the same accuracies.

In regard to the “SVM tuning parameters combined with feature selection” experiment, we used 10 agents and 100 iterations for all optimization techniques. Table 16 displays the accuracy results. As the reader can observe, SSO has outperformed PSO and NGHS for “Industrial” dataset.

Roughly speaking, SSO has obtained very close results to the ones obtained by PSO, being sometimes better than it for “Industrial” dataset considering the three different experiments.

## 5. Conclusions

A considerable number of works have dealt with the problem of estimating SVM parameters, but only a few of them have employed evolutionary-based optimization techniques for this purpose. Very recently, a new optimization technique called Social-Spider Optimization was proposed based on the social dynamics of spiders considering both sexes, i.e., male and female, which one with different functions and responsibilities. SSO has obtained very promising results in several applications, and as far as we know, it has never been applied to the context of this paper: feature selection and SVM parameters tuning. Therefore, the main goal of this paper is to provide an SSO-based framework to improve SVM classification results by means of feature selection and parameters optimization.

The experimental section comprised 10 public datasets with two different kernels (RBF and Polynomial), as well as three different approaches: “feature selection”, “SVM parameters tuning” and “SVM parameters tuning combined with feature selection”. SSO has been compared against with PSO and NGHS, being more accurate (sometimes similar) than them in 5 out of 10 datasets. In regard to the computational load, NGHS was the fastest approach, followed by SSO and PSO.

We have also validated the proposed SSO-SVM in the context of non-technical (commercial) losses detection in power distribution systems, using two private datasets from a Brazilian electrical power company containing commercial and industrial profiles. The results demonstrated all techniques have achieved close results, being PSO the most accurate one for “Commercial” dataset, and SSO for “Industrial” dataset. If we consider the trade-off effectiveness/efficiency, we can consider SSO the preferable approach between PSO and NGHS.

## Acknowledgments

The authors are grateful to FAPESP Grants #2012/06472-9, #2013/20387-7 and #2014/16250-9, as well as CNPq Grants #303182/2011-3, #470571/2013-6 and #306166/2014-3.

## References

- [1] Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20(3):273–97.
- [2] Friedrichs F, Igel C. Evolutionary tuning of multiple SVM parameters. *Neurocomputing* 2005;64:107–17.
- [3] Howley T, Madden M. The genetic kernel support vector machine: description and evaluation. *Artif Intell Rev* 2005;24(3–4):379–95.
- [4] Lessmann S, Stahlbock R, Crone S. Genetic algorithms for support vector machine model selection. In: *Proceedings of the international joint conference on neural networks*; 2006. p. 3063–9.
- [5] Liu Y, Wang G, Chen H, Dong H, Zhu X, Wang S. An improved particle swarm optimization for feature selection. *J Bionic Eng* 2011;8(2):191–200.
- [6] Melgani F, Bazi Y. Classification of electrocardiogram signals with support vector machines and particle swarm optimization. *IEEE Trans Inf Technol Biomed* 2008;12(5):667–77.
- [7] Pereira L, Papa J, de Souza A. Harmony search applied for support vector machines training optimization. In: *Proceedings of the IEEE international conference on computer as a tool*; 2013. p. 998–1002.
- [8] Cawley GC. Model selection for support vector machines via adaptive step-size tabu search. In: *Artificial neural nets and genetic algorithms*. Vienna: Springer; 2001. p. 434–7.
- [9] Cuevas E, Cienfuegos M, Zaldívar D, Pérez-Cisneros M. A swarm optimization algorithm inspired in the behavior of the social-spider. *Expert Syst Appl* 2013;40(16):6374–84.
- [10] Pereira DR, Pazoti MA, Pereira LAM, Papa JP. A social-spider optimization approach for support vector machines parameters tuning. In: *Proceedings of IEEE symposium series on computational intelligence*; 2014. p. 1–6.
- [11] Zou D, Gao L, Wu J, Li S. Novel global harmony search algorithm for unconstrained problems. *Neurocomputing* 2010;73(216–218):3308–18.
- [12] Harris T, Hardin JW. Exact wilcoxon signed-rank and wilcoxon mann-whitney ranksum tests. *Stata J*. 2013;13(2):337–43 (7).
- [13] Zou D, Gao L, Wu J, Li S. Novel global harmony search algorithm for unconstrained problems. *Neurocomputing* 2010b;73(16–18):3308–18.
- [14] Chang C.C., Lin C.J.. LIBSVM: a library for support vector machines; 2001. Software available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [15] Ramos C, Souza A, Chiachia G, Falcão A, Papa J. A novel algorithm for feature selection using harmony search and its application for non-technical losses detection. *Comput Electr Eng* 2011;37(6):886–94.

**Danillo R. Pereira** received his B.Sc. in Computer Science from the São Paulo State University, Brazil in 2006. In 2009, he received his M.Sc. in Computer Science from University of Campinas, Brazil. In 2013, he received his Ph.D. in Computer Science from the University of Campinas, Brazil. He is post-doctoral student at Computer Science Department, São Paulo State University.

**Mario A. Pazoti** received his B.Sc. in Computer Science from the Universidade do Oeste Paulista (UNOESTE). In 2005 he received his M.Sc. in Computer Science from the University of São Paulo, São Carlos, Brazil. He has been Professor at Universidade do Oeste Paulista since 2010. His interests include computer vision and machine learning.

**Luís A.M. Pereira** received his B.S. in Information Systems (2011) and M.S. in Computer Science (2014) from the São Paulo State University, Brazil. He is currently pursuing a Doctorate degree in Computer Science at the Institute of Computing, University of Campinas, São Paulo, Brazil. His research interests include machine learning, pattern recognition and computer vision.

**Douglas Rodrigues** received his B.S. in Informatics (2009) and M.S. in Computer Science (2014) from the São Paulo State University, Brazil. He is currently pursuing a Doctorate degree in Computer Science at the Federal University of São Carlos, São Carlos, Brazil. His research interests include machine learning and meta-heuristic-based optimization.

**Caio O. Ramos** received his B.Sc. (2006) and M.Sc. (2010) in Electrical Engineering from the Univ Estadual Paulista (UNESP), Brazil. In 2014, he received his Ph.D. in Electrical Engineering from the University of São Paulo, Brazil. Currently, he is working as post-doctorate researcher at UNESP. His interests include intelligent systems, protection and power quality.

**André N. de Souza** received his B.Sc. in Electrical Engineering from the Mackenzie University, Brazil. In 1995 and 1999, respectively, he received his M.Sc. and Ph.D. in Electrical Engineering from the University of São Paulo, Brazil. He has been Professor at the Electrical Engineering Department, Univ Estadual Paulista (UNESP), since 2005. His interests include intelligent systems, protection and power quality.

**João P. Papa** received his B.Sc. in Information Systems from the Univ Estadual Paulista (UNESP). In 2005 and 2008, respectively, he received his M.Sc. and Ph.D. in Computer Science from the Federal University of São Carlos and University of Campinas, Brazil. He has been Assistant Professor at the Computer Science Department, Univ Estadual Paulista (UNESP), since 2009. His interests include computer vision and machine learning.