ISSN 2007-9737

WOA-SVM: Whale Optimization Algorithm and Support Vector Machine for Hyperspectral Band Selection and 2D Images Feature Selection

Seyyid Ahmed Medjahed[1,*], Fatima Boukhatem[2]

[1] University of Relizane, Algeria

[2] University of Djilali Liabes, Sidi Belabes, Algeria

seyyidahmed.medjahed@univ-relizane.dz, fatima.boukhatem@univ-sba.dz

Abstract. This paper proposes a new optimization based framework for feature selection and parameters determination of support vector machine, called WOA-SVM and it is applied on band selection in hyperspectral images and feature selection in 2D images. The proposed approach WOA-SVM is based on Whale Optimization Algorithm (WOA), which is a meta-heuristic inspired from the social behaviors of humpback whale and never been benchmarked in the context of feature selection nor parameters determination. A new fitness function is designed. WOA-SVM is tested with three hyperspectral images widely used for band selection and classification. Note that one of the problems in hyperspectral image classification research is the identification of informative bands (band selection). In addition, we demonstrate the efficiency of the proposed approach on Mammographic Image dataset (MIAS). The experimental results prove that the proposed approach is high performance and very competitive approach. The WOA-SVM approach is useful for parameter determination and feature/band selection in SVM.

Keywords. Support vector machine, whale optimization algorithm, cancer diagnosis, parameters determination, feature selection.

1 Introduction

Recently, feature selection has been a primordial step for classification in many domains (images classification, DNA microarray for cancer diagnosis, hyperspectral images classification, etc.). It aims to select the informative and relevant features from a dataset by removing the redundant and irrelevant features [3, 2]. In 2D images classification, feature selection represents the step after the feature extraction and is used to select the optimal subset of features to improve the classification accuracy rate [27, 23].

In the last years, hyperspectral image classification has been an important research field in various applications. The main aim of hyperspectral image classification is to classify each pixel into a specific label. We recall that the hyperspectral image consists of hundreds of spectral bands and each pixel of the image is acquired in light intensity and a large number of spectral bands. Unfortunately, hyperspectral image classification is a very difficult task. The difficulty resides in the fact that we have a reduced number of labeled pixels versus a large number of bands, which cause the Hughes phenomenon.

Hughes phenomenon is a major problem in hyperspectral images classification. This problem is known as the curse of dimensionality caused by the highly correlated and irrelevant bands.

Band selection is a primordial step in hyperspectral image classification. It attempts to select the relevant band by removing the highly correlated and the irrelevant bands without affecting the accuracy of the classification. In other terms, the goal is to select the optimal subset of bands that represents the informative subset and allows achieving a high classification accuracy rate [9, 11]. Several works have been done in the context of the hyperspectral image classification and band selection. In [15], the authors proposed a new band selection approach based on gray wolf optimizer, which is a new optimization approach. The experimentation is conducted on three hyperspectral images, Salinas, Indian Pines and Pavia University. A new objective function is designed based on accuracy and Hausdorff distance. Y. Yuan *et al.* [32] proposed an approach for band selection by using an evolutionary strategy to handle the high computational burden associated with groupwise-selection-based method. S. A. Medjahed *et al.* [17] proposed an approach for band selection based on Jeffries-Matusita distance and a post classification step called CEC (Classification Errors Correction).

This approach was tested on five images, Salinas, Indian Pines, Pavia University, Kennedy Space Center and Botswana. K. Sun *et al.* [29] developed a method namely Minimum Noise Band Selection used for band selection and based on the determination of the quality of each band by using a high SNR and low correlation. In [25] a split-and-merge strategy is proposed for band selection. The aim of this approach is based on two steps. The first one is to split the adjacent bands, and the second one is to merge the highly correlated bands. In [28], the authors developed an approach by using firefly algorithm and optimized extreme learning machine. The goal is to minimize the complexity of the extreme learning machine network using the firefly algorithm.

One of the most robust methods used for hyperspectral image classification and band selection is Support Vector Machine (SVM). Recently, SVM has proven its performance in the classification of hyperspcetral image because its learning ability and generalization capacity by using a few samples of data.

Support vector machine (SVM) is a popular kernel-based learning method and it is widely used for supervised classification [18]. The basic idea of SVM is to find the optimal hypreplane that separates the data by maximizing the margin between two classes. Therefore, the accuracy rate and the quality of classification model generated by the SVM depend largely of certain parameters such as the regularization parameter and kernel parameters. In C-SVM, the parameter C controls the trade-off between errors of the SVM on training data and margin maximization [8]. This parameter must be perfectly adjusted to obtain a satisfactory classification results. A large value of C, we favorite a high penalty for nonseparable points and we may store many support vectors and overfit. A small value of C, produced an underfitting [4]. In addition, the choice of the kernel function and its parameter affect the classification results. If we use the Gaussian kernel, the parameter σ must be set appropriately because a large value leads to overfitting and a small value results in under-fitting. In summary, to classify the dataset using SVM, we must first determine the penalty parameter C, choose a kernel function and set the parameters of the kernel parameter [12].

Another crucial problem when using SVM is how to select the relevant or optimal input subset of features. The quality and the number of features influence directly the classification model and the accuracy rate. The selection of relevant features in the input data plays an important role in the classification. We note that not all the features are informative, some features are irrelevant and redundant, which reduce the accuracy rate. It exists many optimal feature subset and it is possible to achieve the same classification accuracy rate using different feature subset because if two features are correlated a feature can be replaced by other.

In this paper, we propose a novel approach for feature/band selection and parameter determination of support vector machine together. This approach is applied in the context of hyperspectral band selection and 2D images feature selection. The aim is to select the smallest subset of bands/features and to adjust perfectly the SVM parameter and the kernel function parameter.

The proposed approach is called WOA-SVM and it is based on Whale Optimization Algorithm (WOA) which is a new meta-heuristic inspired from the social behavior of humpback whale and recently developed. A binary version of WOA algorithm is also proposed to deal the problem of band/feature selection and a new objective function is designed.

To evaluate the performance of the proposed approach, we have considered three hyperspectral images widely used in the literature $i.e.$ Salinsa, Indian Pines and Pavia University. To demonstrate that the proposed approach can be applied to several type of dataset and not only hyperspectral images, we propose also, to test the proposed approach on the Mammographic Image Analysis Society (MIAS) dataset, and, a complete process based on image feature extraction is proposed.

The rest of paper is structured as follows: In Section 2, an overview of SVM is drawn. Section 3 describes the proposed approach. Section 4 details and discusses the experimental results generated by our approach. Finally, in section 4, the conclusion is drawn with some perspectives.

2 Overview of Support Vector Machines

Support vector machine (SVM) is a classification technique that will construct a separating hyperplane in the attribute space which maximizes the margin between the instances of different classes. Introduced by Vladimir N. Vapnik in 1995 [1, 26], SVM it becomes rather popular since. The models of SVM are closely related to Neural Networks. SVM works well in practice and can be applied to a wide variety of domains such as: bioinformatics, pattern recognition, etc.

The goal of SVM is to find the optimal hyperplane which separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane [30]. The vectors near the hyperplane are called the support vectors.

Find the optimal hyperplane is equivalent to reformulate the classification problem to an optimization problem.

We consider a problem of binary classification. We have an input space $X \subseteq \mathbb{R}^d$, where $d \in \mathbb{N}$, an output space $Y = \{-1, +1\}$, and training set, D where:

$$D = \{(x_1, y_1), ..., (x_N, y_N)\} \in (X \times Y), \quad (1)$$

with hyperplane:

$$f(x) = \langle w, x \rangle + b,\ f(x) = 0. \quad (2)$$

The equation (2) represents the decision rule that, given an input element $x \in X$, if $f(x) \geq 0$ belongs to -1 class; else, x belongs to $+1$ class.

The classifictaion problem consists in determining the hyperplane (2). The modeling of this hyperplane is managed by w and b parameters. These important w and b parameters are learned by the algorithm from the tranining set data.

Each observation x_i is assigned to the class corresponding to the sign of $f(x)$:

$$\begin{array}{ll} \langle w, x \rangle + b \geq +1 & \text{pour } y_i = +1, \\ \langle w, x \rangle + b \leq -1 & \text{pour } y_i = -1. \end{array} \quad (3)$$

Which can be combined into:

$$y_i(\langle w, x_i \rangle + b) - 1 \geq 0\ \forall i. \quad (4)$$

The optimal separation hyperplane represents the best choice to divide the input element of X into the output classes Y. Since the optimal separation hyperplane maximize the $\frac{2}{\|w\|}$ margin and it is equivalent to minimize $\frac{1}{2}\|w\|^2$. Finding this optimal hyperplane is equivalent to solving the following optimization problem:

$$\begin{array}{ll} \min_{w,b} & \frac{1}{2}\|w\|^2, \\ s.\ c. & y_i(\langle w, x_i \rangle + b) \geq 1, \\ & i \in \{1, ..., N\}. \end{array} \quad (5)$$

The problem (5) is the primal form of quadratic optimization problem which is difficult to solve when the dataset is very large. Therefore, it is very interesting to transform it to a dually form by introducing the Lagrange multiplier and using the Karush–Kuhn–Tucker condition (we compute the derivatives respect with b and w variables):

$$\begin{array}{ll} \min_\alpha & -\sum_{i=1}^{N}\alpha_i+\frac{1}{2}\sum_{i,j}y_iy_j\alpha_i\alpha_j\left\langle x_i,x_j\right\rangle, \\ s.\ c. & \sum_{i=1}^{N}\alpha_iy_i=0, \\ & \forall i\in\{1,...,N\},\ \alpha_i\geq 0. \end{array} \tag{6}$$

By solving the equation (6) we determine the Lagrange multipliers α_i^* and the optimal hyperplane is given by:

$$\begin{aligned} w^* &= \sum_{i=1}^{N}\alpha_iy_ix_i, \\ b^* &= -\frac{1}{2}\left\langle w^*,x_r+x_s\right\rangle, \end{aligned} \tag{7}$$

$$H(x)=sign(\langle w^*,x\rangle+b^*). \tag{8}$$

where x_r and x_s are any support vector from each class satisfying :

$$\alpha_r,\alpha_s>0,\ y_r=-1,y_s=1. \tag{9}$$

Finally, by using Wolfe Dual Form and Kuhn Tucker conditions, the equation (8) can be written as:

$$H(x)=sign(\sum_{i=1}^{N}\alpha_s^*y_s(x\cdot x_s)+b).$$

Solving the optimization problem requires the use of optimization quadratic algorithms such as: Sequential Minimal Optimization (SMO), Trust Region, Interior Point, Active-Set, etc. In this study, the SMO method is used [6, 22, 31].

Generally, the data are not linearly separable (the case of presence of error points inside the two classes to be classified), in order to extend the SVM methodology to handle data that is not fully linear separable, the quadratic optimization problem must be relaxed by introducing positive slack variables to allow for misclassified points:

$$\begin{array}{ll} \min_{w,b,\xi} & \frac{1}{2}\left\|w\right\|^2+C\sum_{i=1}^{N}\xi_i, \\ s.\ c. & y_i(\langle w,x_i\rangle+b)\geq 1-\xi_i, \\ & \xi_i\geq 0, \\ & i\in\{1,...,N\}. \end{array} \tag{10}$$

The slack variables represent the error distance of the wrong points from the optimal separating hyperplane : $\xi=(\xi_1,\xi_2,...,\xi_N)$, where $\forall_i\ \xi_i\geq 0$.

The parameter C controls the trade-off between the slack variable penalty and the size of the margin. By using the Lagrange multipliers, this optimization formulation can be transformed into the following dual problem:

$$\begin{array}{ll} \min_\alpha & -\sum_{i=1}^{N}\alpha_i+\frac{1}{2}\sum_{i,j}y_iy_j\alpha_i\alpha_j\left\langle x_i,x_j\right\rangle, \\ s.\ c. & \sum_{i=1}^{N}\alpha_iy_i=0, \\ & \forall i\in\{1,...,N\},\ 0\leq\alpha_i\leq C. \end{array} \tag{11}$$

In the dual problem, the parameter C is an upper bound for α_i.

In the cases where the data cannot be classified explicitly in the current dimensional space, the basic idea of the SVM is to map input data into a convenient space where they can be separated with an opportune hyperplane by using the kernel function:

$$\begin{array}{lll} K: & \mathbb{R}^n\times\mathbb{R}^n & \to\mathbb{R} \\ & (x_i,x_j) & \to\left\langle\Phi(x_i),\Phi(x_j)\right\rangle. \end{array} \tag{12}$$

With kernel function, the quadratic program can be written again this way:

$$\begin{array}{ll} \min_\alpha & -\sum_{i=1}^{N}\alpha_i+\frac{1}{2}\sum_{i,j}y_iy_j\alpha_i\alpha_jK(x_i,x_j) \\ s.\ c. & \sum_{i=1}^{N}\alpha_iy_i=0 \\ & \forall i\in\{1,...,N\},\ \alpha_i\geq 0. \end{array} \tag{13}$$

Kernel functions are a very powerful tool to solve classification problem. Several researches have been devoted to constructing a kernel more exotic and adapted to a special problem by respecting the Mercer's theorem. The most frequently used such functions are:

Linear $K(x_i,x_j)=(x_i\cdot x_j)$

Polynomial $K(x_i,x_j)=[(x_i\cdot x_j)+1]^d$ where $d\in\mathbb{N},d\neq 0$

Gaussian $K(x_i,x_j)=\frac{-\|x_i-x_j\|^2}{2\sigma^2}$, also called radial basis function (RBF).

3 Proposed Approach WOA-SVM

3.1 Problem Formulation

In this study, the parameter determination of SVM and the feature selection is designed as a combinatorial optimization problem. The problem is modeled as follows.

Let $F = \{F_1, \ldots, F_n\}$ a set of features. Let $X = \{X_1, ..., X_n\}$ a binary vector with:

$$X_i = \begin{cases} 1 \text{ if } F_i \text{ is selected,} \\ 0 \text{ otherwise.} \end{cases} \tag{14}$$

For SVM and kernel function parameters, we determine the upper and lower bound:

$$C \in [l_c, u_c],$$

$$\gamma \in [l_\gamma, u_\gamma].$$

The decision variable has the following form:

$$X = [X_1, ..., X_n],\ [C, \gamma].$$

The basic idea is to find the optimal subset of features and the optimal parameter C and γ that provided a high classification accuracy rate. Therefore, the objective function is to maximize the accuracy rate in classifying the testing data by using SVM. This is equivalent to an optimization problem [10]. We propose to use the Whale Optimization Algorithm to solve this optimization problem.

3.2 Whale Optimization Algorithm

Whale optimization algorithm (WOA) is a new meta-heuristic optimization algorithm developed by Ali Mirjili [19]. It is mimicking the hunting behaviour of humpback whales. WOA is slightly similar to Gray Wolf Optimizer recently developed by Seyedali Mirjalili [20] and used for feature selection [16]. The difference resides in the simulation hunting behavior and the use of bubble-net attacking of humpback whales.

The process of humpback whales hunting represents the optimization process and is divided into three phases:

3.2.1 Encircling Prey

Humpback whales detect the position of the prey and encircle them. In the optimization process, the prey is the optimal solution and it is not known a priori, WOA algorithm assumes that the current best candidate solution the target prey or is close to the optimum [19]. The position of search agents are updated according to the position of the best search agent. The encircling behaviors is given as follows:

$$\overrightarrow{D} = \left|\overrightarrow{C}.\overrightarrow{X^*}(t) - \overrightarrow{X}(t)\right|, \tag{15}$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X^*}(t) - \overrightarrow{A}.\overrightarrow{D}, \tag{16}$$

where t is the current iteration, $\overrightarrow{A}$ and $\overrightarrow{C}$ are coefficient vectors, X^* is the position vector of the best solution [19]. The vectors $\overrightarrow{A}$ and $\overrightarrow{C}$ are computed as follows:

$$\overrightarrow{A} = 2\overrightarrow{a}.\overrightarrow{r} - \overrightarrow{a}, \tag{17}$$

$$\overrightarrow{C} = 2.\overrightarrow{r}, \tag{18}$$

where $\overrightarrow{a}$ is linearly decresed from 2 to 0 over the iterations and $\overrightarrow{r}$ is a random vector in 0 and 1.

3.2.2 Bubble-net Attacking

This phase represents the exploitation phase and it consists of two approaches: the first one is the shrinking encircling mechanism and is reached by decreasing the value of $\overrightarrow{a}$, and also, $\overrightarrow{A}$ decreased. So, $\left|\overrightarrow{A}\right| \leq 1$, the new position of search agent is defined between the original position on the agent and the position of the current best agent [19].

The second approach is the spiral updating position and it is based on the computing of the distance between humpback whale and at (X, Y) and prely located (X^*, Y^*) [19]. To imitate the helix-shaped mouvement of humpback whale, we must use the following equation (19):

$$\overrightarrow{X}(t+1) = \overrightarrow{D'}.e^{bl}.\cos(2\pi l) + \overrightarrow{X^*}(t), \tag{19}$$

where $\overrightarrow{D'} = \left|\overrightarrow{X^*}(t) - \overrightarrow{X}(t)\right|$ indicates the distance between the ith whale to the prey. b is a constant for defining the shape of logarithmic spiral and l is random number between -1 and 1. The mathematical model of updating spiral position is as follows:

$$\vec{x}(t+1) = \begin{cases} \overrightarrow{X^*}(t) - \vec{A}.\vec{D} & if\ p < 0.5, \\ \overrightarrow{D'}.e^{bl}.\cos(2\pi l) + \overrightarrow{X^*}(t) & if\ p \geq 0.5, \end{cases} \tag{20}$$

where p is random variable between 0 and 1.

As seen, the equation (20) simulates the shrinking circle around the prey and the spiral-shaped path. We assume that there is a probability of 50% to choose between the shrinking encircling or the spiral model to update the position of the whales during the optimization [19].

3.2.3 Search for Prey

This phase is the exploration phase, in other terms, in the approach to search the prey (optimum). Humpback whale search randomly the prey according to the position of randomly chosen search agent instead to the best search agent found so far [19]. We use $\left|\vec{A}\right| > 1$ to force search agent to move far away from a reference whale. This mechanism emphasize exploration and allow to WOA to perform global search [19]. The mathematical model is as follows:

$$\vec{D} = \left|\vec{C}.\overrightarrow{X_{\text{rand}}} - \vec{X}\right|, \tag{21}$$

$$\vec{X}(t+1) = \overrightarrow{X_{\text{rand}}} - \vec{A}.\vec{D}. \tag{22}$$

where $\overrightarrow{X_{\text{rand}}}$ is random position vector (a random whale). A random search agent is chosen when $|A| \geq 1$, while the best solution is selected when $|A| < 1$ for updating the position of the search agents. WOA can be considered a global optimizer because it includes exploration/exploitation ability [19].

The general schema of WOA algorihm can be described as follows [19]:

Algorithm 1 Whale Optimization Algorithm

Initialization
Initialize randomly the search agents position $X(i,j)$ $(i = 1, \ldots, number\ of\ search\ agent, j = 1, \ldots, dimension)$
Compute the fitness of each search agent
X^* is the best solution
Main Loop
for $t = 1$ to $max\ number\ of\ iteration$ **do**
  for $each\ search\ agent\ i$ **do**
    Update a, A, C, l and p
    if $p < 0.5$ **then**
      if $\left|\vec{A}\right| < 1$ **then**
        Update the position of the current search agent using equation (16)
      else
        if $|A| \geq 1$ **then**
          Select a random search agent $\overrightarrow{X_{\text{rand}}}$
          Update the position of the current search agent using equation (22)
        end if
      end if
    else
      if $p \geq 0.5$ **then**
        Update the position of the current search agent using equation (19)
      end if
    end if
  end for
  Check if any search agent goes beyond the search space and amend it
  Compute the fitness of each search agent Update
  Update X^* if there is a better solution
end for
Return X^*

3.3 WOA-SVM Algorithm

The proposed approach WOA-SVM allows determining the optimal SVM parameter and the relevant subset of features, which provide a high classification accuracy rate. We slightly modified the WOA algorithm to combine it with SVM and to take into consideration the problem of parameter determination and feature selection. This is why,

ISSN 2007-9737

a binary version of WOA is proposed to use it for feature selection. The whale position is a defined as binary vector and the problem is to select or no not a feature [13]. If $X_i = 1$, the feature is selected and used for classification, else, if $X_i = 0$, the feature is removed. To take account a binary variable for feature selection, we propse to use the sigmoid function as follows:

$$\begin{aligned} S(X_i) &= \frac{1}{1+\exp(-X_i)}, \\ X_i' &= \begin{cases} 1 \; if \; S(X_i) \geq 0.5, \\ 0 \; if \; S(X_i) < 0.5. \end{cases} \end{aligned} \tag{23}$$

WOA-SVM algorithm for feature selection and parameter determination is described in the following algorithm: (SEE LAST PAGE).

Lines $1-8$ represents the initialization of parameters algorithm. Lines $3-5$, we split the dataset into three subsets: training set (T_1), testing set (T_2) and the validation set (T_3) and in each iteration, we generate randomly a training set (T_1') and testing set (T_2') from (T_1) and (T_2). This strategy allows to the algorithm to avoid the problem of over fitting. The objective function is the classification error rate computed by using the SVM classifier. Line 9 is the main loop and it is determined by the maximum number of iteration. Line $10-29$ is the second main used to update the position of each search (whale) according to the different situations. Lines 31 and 32 generate randomly training and testing set. Lines $33-34$ remove the features that have $X'(i,j) = 0$. These features will not be used to build the classification model. Lines $36-37$ compute the objective function and update the better solution X^*. finally, line 39 return the best solution X^*.

Algorithm 2 WOA-SVM Algorithm

1: Initialize randomly the search agents position $X(i,j)$ $(i = 1,\ldots, number\ of\ search\ agent,\ j = 1,\ldots, dimension)$
2: Initialize SVM parameter C and Gaussian kernel parameter σ
3: Split dataset to training set T_1, testing set T_2 and validation set T_3
4: Generate randomly Training set $T_1^{'}$ from T_1 and Test set $T_2^{'}$ from T_2
5: Train SVM classifier over $T_1^{'}$ and evaluate its over $T_2^{'}$
6: Compute the objective function of each search agent
7: X^* is the best solution
8: **for** $t = 1$ to $max\ number\ of\ iteration$ **do**
9:   **for** $i = 1$ to $max\ number\ of\ search\ agent\ i$ **do**
10:     Update a, A, C, l and p
11:     **if** $p < 0.5$ **then**
12:       **if** $\left|\vec{A}\right| < 1$ **then**
13:         Update $X(i,j)$ position of the search agent and $[C,\gamma]$ using equation (16)
14:         Update $X'(i,j)$ using (23)
15:       **else**
16:         **if** $|A| \geq 1$ **then**
17:           Select a random search agent $\overrightarrow{X_{\mathrm{rand}}}$
18:           Update $X(i,j)$ position of the search agent and $[C,\gamma]$ using equation (22)
19:           Update $X'(i,j)$ using (23)
20:         **end if**
21:       **end if**
22:     **else**
23:       **if** $p \geq 0.5$ **then**
24:         Update $X(i,j)$ position of the search agent and $[C,\gamma]$ using equation (19)
25:         Update $X'(i,j)$ using (23)
26:       **end if**
27:     **end if**
28:   **end for**
29:   Check if any search agent goes beyond the search space and amend it
30:   Generate randomly Training set $T_1^{'}$ from T_1 and Test set $T_2^{'}$ from T_2
31:   $T_1^{'} \leftarrow T_1^{'} - \{features\ with\ X'(i,j) = 0\}$
32:   $T_2^{'} \leftarrow T_2^{'} - \{features\ with\ X'(i,j) = 0\}$
33:   Train SVM classifier over $T_1^{'}$ and evaluate its over $T_2^{'}$
34:   Compute the objective function of each search agent
35:   Update X^* and and $[C,\gamma]$ if there is a better solution
36: **end for**
37: Return X^* and and $[C,\gamma]$

3.4 Objective Function

Generally, the fitness or objective function is the error rate or the classification accuracy rate. In [16], the authors proposed a five objective functions based on the measure of discrimination of each features and the classification accuracy rate. The authors have used the Jeffries-Matusita distance and Hausdorff distance to measure the class separability of selected features. In [7], the authors have used the F-score and the accuracy rate to design the fitness function. In [13], the authors have proposed an objective function based on the balance error rate and F-score. In [5], a fitness function composed of two terms is proposed. The first one is the accuracy and the second is the cost of selected of features. In this study, we propose to use two objective functions.

3.4.1 First Objective Function

The first objective function J_1 is the classification accuracy rate. The value of J_1 is between 0 and 1. In this case, we select the subset of features and determine the parameters that provide the high classification accuracy rate. The objective function can be given as follows [16]:

$$\begin{cases} accuracy = \dfrac{\sum_{i=1}^{|N|} assess(x_i)}{|N|}, \\ assess(x_i) = \begin{cases} 1\ if\ classify(x_i) = c \\ 0\ otherwise, \end{cases} \end{cases} \tag{24}$$

where, N represents the instance set, x_i is the instance to be classified, $assess(x_i)$ is the classification function. It sets to 1 if the class is the true label of x_i ($classify(x_i) = c$) and 0 otherwise.

3.4.2 Second Objective Function

The second objective function J_2 is composed of two important terms. The first term of the objective function is the weight vector w of the SVM given as follows:

$$w = \sum_l \alpha_l y_l x_l, \tag{25}$$

$$f_1(X) = \sum_i X_i \times w_{F_i}, \tag{26}$$

where y_l is the label of the instance x_l, α_l is the Lagrange multipliers, X is a binary vector. If $X_i = 0$ then the feature F_i is not selected, else, $X_i = 1$ then the feature F_i is selected, w_{F_i} weight of the selected feature F_i.

The summation in equation (25) is taken over the training set. The weight vector w_{F_i} represents a measure of the ranking of the feature F_i and their influence on the margin of separation of the labels, in other terms, their influence on the separated hyperplane.

The second term of the objective function measures the ability of each feature to regroup the classes. We assure that a feature is a good feature if it regroups perfectly their instance around its gravity center. Therefore, we propose to compute for each class, the distance between their instances and its gravity center. The features that minimize this distance are selected. Figure 1 illustrates this idea. The second term of the objective function is given as follows:

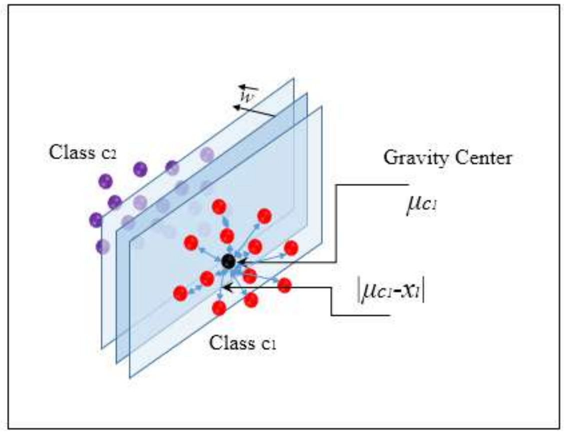


Fig. 1. The proposed measure to compute the ability of feature to regroup the instances of each class

For binary problems, the proposed $d(c_1, c_2)$ distance for two classes c_1 and c_2, is given as follows:

$$d(c_1, c_2) = \sum_l \left|\mu_{c_1} - x_l^{c_1}\right| + \sum_m \left|\mu_{c_2} - x_m^{c_2}\right|, \tag{27}$$

where, μ_{c_1} μ_{c2} is the gravity center (average) of class c_1 and class c_2, $x_l^{c_1}$ instance with $y_l = c_1$, $x_m^{c_2}$ instance with $y_m = c_2$,

For multiclass problems, we defined the proposed distance as follows:

$$d_{F_i} = \frac{1}{p \times (p-1)} \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} d(c_i, c_j), \tag{28}$$

$$f_2(X) = \sum_i X_i \times d_{F_i}, \tag{29}$$

where d_{F_i} is the distance between the gravity center and each instance of each class of the feature F_i, X is a binary vector. If $X_i = 0$ then the feature F_i is not selected, else, $X_i = 1$ then the feature F_i is selected.

The final fitness function is s the weighted sum of the both terms f_1 and f_2:

$$J_2(X) = \theta_1 \times f_2(X) + \theta_2 \times (f_2(X))^{-1}, \tag{30}$$

where θ_1 and θ_2 are weights of f_1 and f_2 respectively. We can balance between the functions and give more importance to f_1 or f_2.

4 Experimental Results

4.1 Datasets

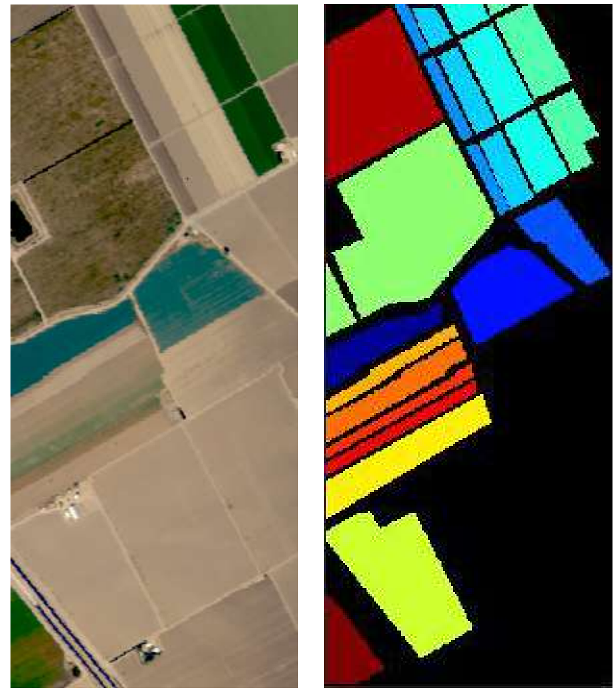
In order to analyze the performance of the proposed approach WOA-SVM, we conduct two experimentations. The first experimentaion is conducted under three hyperspectral images very used in the literature: Salinas, Indian Pines and Pavia University.

4.1.1 Salinas Image

The first hyperspectral image used in this experimentation is knows as the Salinas image, which is acquired by AVIRIS (Airborne Visible InfraRed Imaging Spectrometer) over the Salinas Valley, Southern California, USA. This image is 512×217 pixels and composed of 224 bands in the spectral range 0.4 μm to 2.5 μm. It contains 16 ground truth classes: Broccoli-green-weeds-1, Broccoli-green-weeds-2, Fallow, Fallow-rough-plow, Fallow-smooth, Stubble, Celery, Grapes-untrained, Soil-vinyard-develop, Corn-senesced-green-weeds, Lettuce-romaine-4wk, Lettuce-romaine-5wk, Lettuce-romaine-6wk, Lettuce-romaine-7wk, Vineyard-untrained and Vineyard-vertical-trellis.

4.1.2 Indian Pines Image

The second hyperspectral image is called Indian Pines, which is taken over the agricultural area of Northwestern of Indiana, USA. This image is 145×145 and composed of 220 bands in the spectral range 0.5 μm to 2.5 μm. It was acquired by AVIRIS and it contains 16 ground truth classes: Alfalfa, Corn-notill, Corn-mintill, Corn, Grass-pasture, Grass-trees, Grass-pasture-mowed, Hay-windrowed, Oats, Soybean-notill, Soybean-mintill, Soybean-clean, Wheat, Woods, Buildings-Grass-Trees-Drives, and Stone-Steel-Towers.



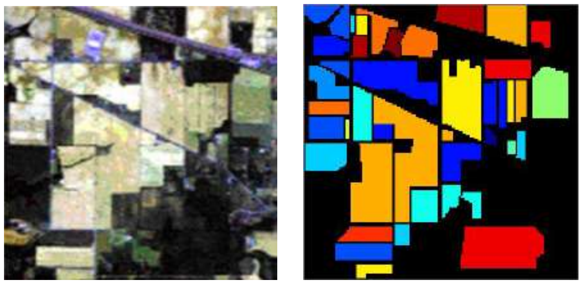
(a) (b)

Fig. 2. Salinas hyperspectral image. (a) Color compose. (b) Ground truth

4.1.3 Pavia University Image

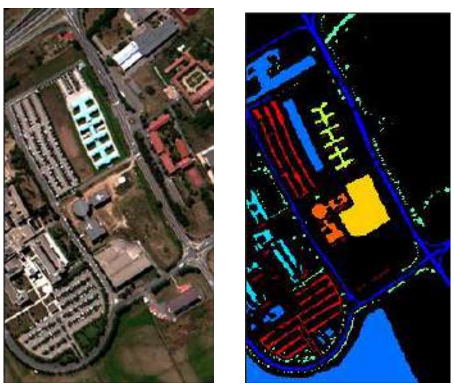
The third hyperspectral image used in this study is the Pavia University image taken over the urban area of Pavia University. This image is 610×340 pixels and it was collected by ROSIS. It composed of 103 bands in the spectral range from $0.4\mu m$ to 0.86 μm. The ground truth differentiates 9 classes: Asphalt, Meadows, Gravel, Trees, Painted Metal Sheets, Bare Soil, Bitumen, Self-Blocking Bricks, and Shadows.

The second experimentation is conducted under mammographic image dataset. We use MIAS dataset (Mammography Image Analysis Society) which is widely known dataset. MIAS dataset contains 322 images classed into three categories: normal, benign and malign. The category normal is a set of mammogram image without breast cancer. The benign and malign is a collection of abnormal mammogram images. MIAS dataset contains 208 normal images, 63 benign images



(a) (b)

Fig. 3. Indian Pine hyperspectral image. (a) Color image. (b) Ground truth



(a) (b)

Fig. 4. Pavia University hyperspectral image. (a) Color Image. (b) Ground truth.

and 51 malign images. The size of each image is 1024×1024 pixels.

To extract the region of interest (ROI) in the image, MIAS dataset provides the coordinates X, Y and the radius in pixels of each abnormal images. Figure 5 shows some MIAS mammogram images.

4.2 Parameters Setting

For each classifier system, the number of samples used for training and testing phases must be determined. For the first experimentation over the hyperspectral images, we split the pixels into two subsets, 10% of the pixels are used for the training phase and the remaining 90% of pixels are used for test phase.

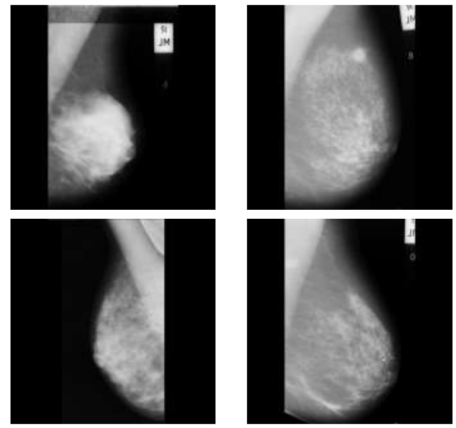


Fig. 5. Some mammogram images taken from MIAS dataset

In the second experimentation, , we propose to split the dataset into three subsets: training set, testing set and validation set. 50% of samples are used for training, 20% of samples are used for the test and the remaining 30% of samples are used for validation.

The parameters of WOA-SVM are adjusted as follows: the number of search agent is 20. The algorithm stops when the number of iterations fixed to 500 is achieved.

The search interval of SVM parameter C and Gaussian kernel parameter σ are:

$$C \in [1, 500],$$

$$\gamma \in [0.01, 50].$$

These parameters have demonstrated their performance and were chosen by experimentation for the both datasets.

4.3 Application in Hyperspectral Images

In this section, we analyze the results obtained by the proposed approach WOA-SVM in term of classification average accuracy rate (AA), classification overall accuracy rate (OA), individual class accuracy rate and the number of selected bands. Table 1 describes the classification average accuracy rate, classification overall accuracy rate and the number of selected bands obtained by WOA-SVM.

ISSN 2007-9737

Table 1. The results obtained by the WOA-SVM approach

Hyperspectral image	Salinas		Indian Pines		Pavia University	
Objective function	J_1	J_2	J_1	J_2	J_1	J_2
AA	95,96	96,22	72,41	71,87	87,77	87,16
OA	91,35	91,85	72,60	74,61	89,24	88,72
Number of selected bands	148	112	120	98	51	43

Table 2. Classification accuracy rate obtained by WOA-SVM over Salinas scene and compared with previous works

Class	Feature Selection Approaches								This Study	
	mRmR	cmim	Relief	GA	PSO	GSA	BBA		J_1	J_2
Brocoli_green_weeds_1	93,97	98,69	91,98	98,26	99,17	99,75	98,67	98,32	99,42	99,50
Brocoli_green_weeds_2	95,84	99,56	61,86	99,60	99,55	99,77	99,41	99,60	99,73	99,96
Fallow	85,39	98,73	94,18	98,04	99,40	99,57	98,73	99,30	99,16	99,41
Fallow_rough_plow	99,73	99,73	89,25	99,73	99,40	99,28	99,40	99,73	99,40	99,16
Fallow_smooth	94,91	96,13	97,34	95,01	98,50	98,13	98,31	96,45	97,70	98,94
Stubble	99,27	99,81	98,39	99,94	99,91	99,87	99,95	99,94	99,96	99,92
Celery	95,98	99,44	61,03	99,37	99,72	99,67	99,58	99,37	99,67	99,77
Grapes_untrained	76,54	82,54	77,23	79,34	81,70	81,42	79,58	83,25	80,08	82,02
Soil_vinyard_develop	97,76	97,92	98,57	99,38	99,48	99,57	99,62	99,27	99,81	99,65
Corn_green_weeds	75,60	91,80	78,42	91,61	96,49	95,57	95,17	91,69	95,63	94,97
Lettuce_romaine_4wk	60,94	93,33	92,05	92,05	99,84	98,43	98,90	94,04	98,28	99,06
Lettuce_romaine_5wk	86,64	99,74	92,54	98,83	99,56	99,91	100	99,87	100	99,91
Lettuce_romaine_6wk	96,45	98,23	96,45	97,82	98,72	99,45	98,54	97,82	99,64	98,91
Lettuce_romaine_7wk	92,41	94,39	92,64	93,57	97,97	96,41	97,97	94,16	96,42	97,04
Vinyard_untrained	57,33	60,00	47,52	57,88	70,90	72,85	67,75	63,54	71,29	71,68
Vinyard_vertical_trellis	76,42	98,27	82,16	97,86	99,07	99,63	98,15	98,06	99,17	99,63
AA	86,57	94,27	84,48	93,64	96,21	96,20	95,61	94,65	95,96	96,22
OA	83,79	89,55	79,28	88,57	91,70	91,86	90,67	90,35	91,35	91,86

Table 1 contains the AA, OA the number of selected bands obtained by the proposed approach WOA-SVM for each hyperspectral image and under the both objective function J_1 and J_2. The first row of table 1 represents the name of the hyperspectral images. The columns $2, 3$ and 4 are the AA, OA and the number of selected bands obtained by the objective function J_1 and J_2.

The results described in table 1 show the efficacy of the proposed approach. As observed in table 1, the outcome of WOA-SVM under the three hyperspectral images was quite appropriate, we record 95.96% and $96,22\%$ of average accuracy in Salinas scene for the both objective functions J_1 and J_2 respectively. For Indian Pines scene, the results are very satisfying, as best accuracy of $72,41\%$ and $71,87\%$ as average accuracy. The performance of Pavia University scene shows that the proposed approach attaining $87,77\%$ and $87,16\%$ for J_1 and J_2 respectively as maximum average accuracy rate. The analysis of the results obtained by WOA-SVM in terms of number of selected bands is better and we remark and advantage of the results using J_2 compared to J_1.

To demonstrate the performance of the proposed approach WOA-SVM, we compare our approach to several feature selection approaches defined in the literature.

Tables 2, 3 and 4 compare the AA, OA and individual class accuracy rate between WOA-SVM

ISSN 2007-9737

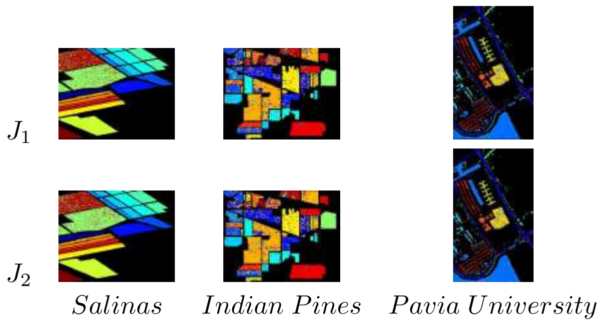


Fig. 6. The classification maps obtained by our approach for the five objective functions and applied to Pavia University, Indian Pines and Salinas hyperspectral data sets

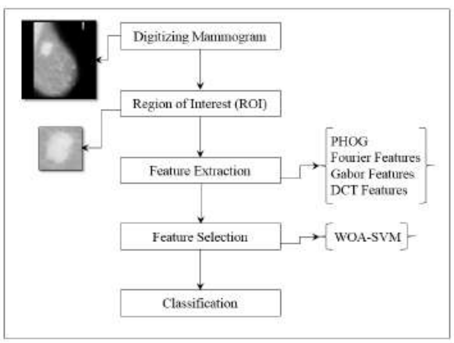


Fig. 7. Framework of the proposed process

and other approaches for each hyperspectral image. The results is reported from [15], [17], [14].

Tables 2, 3 and 4 shows the comparison of the AA, OA and individual class accuracy obtained by WOA-SVM and some previous works. The AA, OA and individual class accuracy rate for Salinas scene, Indian Pines scene and Pavia University scene respectively. The first column of tables represents the name of classes and the last two rows represent the AA and OA. The second column represents the band selection approaches used for comparison and it contains seven sub columns each one represents a band selection approach. The third column represents the results obtained by using all the bands under two classifiers: k-nearest neighbor and SVM. The last column represents the results obtained by our study for the both objective function J_1 and J_2.

For the comparison, we use seven filter approaches: mRMR (Max-Relevance Min-Redundancy), CMIM (Conditional Mutual Info Maximisation), and Relief. Also, we compare the proposed approach to four wrapper approaches: PSO (Particle Swarm Optimization [**?**]), BGSA (Binary Gravitational Search Algorithm [24]), BBA (Binary Bat Algorithm [21]) and GA (Genetic Algorithm [5]). In addition, we propose to compare our approach to K-nearest neighbor and Support Vector machines classifiers by using all the features (without bands selection).

As shows in tables 2, 3 and 4, the results obtained by WOA-SVM using the objective function J_1 and J_2 perform significantly better than the other approach with a slight advantage to J_2 in Salinas and Indian Pines hyperspectral images. All the wrapper approaches have provided good results compared to filter approaches; this is because wrapper approaches make a repetitive call to a classifier system. Wrapper approaches use accuracy rate as an objective function and attempt to optimize this last contrary to filter approaches, which are independent to the classifier system.

By analyzing the experimental results, we conclude that WOA-SVM performs well and provides satisfying results compared to several feature selection approach.

The visual results (classification maps) is illustrated in figure 6.

Figure 6 show the classification map generated by the proposed approach WOA-SVM for the both objective function.

4.4 Application in Mammography Image Dataset

In this section, we analyze the performance of the proposed approach WOA-SVM in MIAS dataset. Figure 7 illustrates the proposed process to classify mammogram image.

The process of mammogram image classification contains two important steps: feature extraction and feature selection. Feature extraction step consists to extract the feature of each image. In this study, we propose to combine four types of features obtained by four feature extraction methods:

ISSN 2007-9737

Table 3. Classification accuracy rate obtained by WOA-SVM over Indian Pines scene and compared with previous works

Class	Feature Selection Approaches									This Study	
	mRmR	cmim	Relief	GA	PSO	GSA	BBA		knn	J_1	J_2
Alfalfa	0,00	16,22	0,00	0,00	60,71	50,57	60,71	84,44	21,63	60,70	60,71
Corn-on till	31,67	55,12	54,42	52,93	60,91	69,89	61,61	36,79	52,60	56,83	62,54
Corn-min till	18,37	47,89	24,10	28,16	64,65	69,67	56,42	40,67	51,21	59,44	60,04
Corn	1,58	40,53	8,42	14,74	46,15	46,85	46,15	72,31	31,06	43,36	42,66
Grass/pasture	70,28	83,46	57,11	82,17	88,62	88,96	89,31	80,40	85,28	87,59	85,86
Grass/tree	66,95	93,15	86,46	94,35	93,15	95,89	94,97	78,93	96,91	93,61	96,12
Grass/pasture-mowed	4,35	86,96	0,00	17,39	70,23	88,23	52,94	95,38	78,27	88,24	88,24
Hay-windrowed	51,44	96,08	99,22	96,87	98,60	97,90	95,47	76,11	98,96	94,08	98,26
Oats	0,00	6,25	0,00	0,00	50,00	50,00	50,00	100	0	58,33	33,33
Soybeans-no till	13,24	66,84	41,65	59,38	70,68	78,93	66,95	53,80	70,83	71,75	72,77
Soybeans-min till	62,07	75,51	67,92	74,13	77,66	82,41	75,49	39,73	77,75	73,93	76,65
Soybeans-clean till	11,58	45,47	12,63	32,00	59,26	58,98	58,42	49,12	44,43	52,53	53,65
Wheat	66,46	92,07	91,46	94,51	90,74	91,86	95,12	91,42	96,35	91,87	93,50
Woods	92,98	94,76	90,51	92,98	91,17	42,38	92,35	81,31	94,97	90,38	91,83
Bldg-grass-tree-drives	10,68	11,33	7,44	11,65	48,70	46,55	48,70	47,05	18,13	44,83	42,67
Stone-steel towers	18,67	65,33	25,33	73,33	92,85	89,28	87,50	94,11	81,34	91,07	91,07
AA	32,52	61,06	41,68	51,54	72,75	71,77	70,76	70,10	62,49	72,41	71,87
OA	46,59	69,48	57,67	64,86	74,55	74,03	73,89	56,42	71,01	72,60	74,61

Table 4. Classification accuracy rate obtained by WOA-SVM over Pavia University scene and compared with previous works

Class	Feature Selection Approaches									This Study	
	mRmR	cmim	Relief	GA	PSO	GSA	BBA			J_1	J_2
Asphalt	83,39	79,89	83,52	82,28	89,21	90,44	89,77	84,93	86,53	89,52	89,67
Meadows	95,82	93,17	94,18	96,91	90,98	91,64	92,62	70,79	97,93	95,48	95,08
Gravel	53,21	53,93	47,38	43,75	72,85	77,38	73,01	67,16	70,00	74,60	73,65
Trees	77,20	60,03	53,47	79,73	88,52	87,87	85,42	97,77	83,12	89,83	89,40
Painted Metal Sheets	99,54	98,88	89,78	99,16	99,75	94,25	96,75	99,46	99,08	99,75	99,26
Bare Soil	48,78	26,47	62,72	58,25	70,45	70,70	71,00	92,83	58,83	71,90	70,21
Bitumen	81,11	78,67	75,75	75,47	80,59	84,96	83,20	90,42	85,62	85,46	84,21
Self-Bloking Bricks	85,71	81,74	79,97	83,47	81,70	80,02	80,35	92,78	87,58	83,39	82,99
Shadows	100	75,33	100	100	98,01	99,82	97,82	98,11	99,87	100	100
AA	80,53	71,98	76,31	79,89	85,78	86,12	85,77	88,25	85,40	87,77	87,16
OA	83,82	77,29	81,81	84,57	87,38	87,97	87,28	81,01	97,85	89,24	88,72

— Features from F_1 to F_{680} are obtained by Pyramid Histogram of Oriented Gradients,

— Features from F_{681} to F_{688} are obtained by Fourier,

— Features from F_{698} to F_{755} are obtained by Gabor,

— Features from F_{756} to F_{759} are obtained by DCT.

The total number of features is 759 and not all the features can be used for the classification.

The second step of the process is to select from 759 features the most informative and relevant features to provide a high classification accuracy rate and a good classification model. We use the proposed approach WOA-SVM to select the optimal subset of features. Table 5 and table 6 table the results obtained by WOA-SVM compared to other feature selection approaches.

From Table 5 and table 6, we clearly observed that WOA-SVM achieved a high accuracy rate. The maximum of accuracy obtained under the validation test is 100%, the mean accuracy rate is 98.82% obtained by the objective function J_1 and 98.71% by J_2. A small advantage is observed for J_2 with a minimum of accuracy 96.61% compared to the minimum of accuracy provided by J_2 (94.91%).

The experimental results demonstrated the superiority of WOA-SVM with regards to other approaches.

This paper can be summarized with the following points:

1. The proposed approach is called WOA-SVM and is based on a recent meta heuristic called Whale Optimization Algorithm which mimicking the hunting behavious of humpback whales.

2. The problem of feature/band selection and parameters determination of SVM and kernel function is designed as a combinatorial optimization problem.

3. A binary version of WOA is developed to deal the problem of feature selection.

4. A new fitness function is designed and it is composed of two important terms: the first term is based on the weight vector w generated by SVM which as the ranking criterion. This criterion represents the ability of a feature to affect the margin of separation (optimal hyperplane). The second term computes the ability of a feature to regroup the instances of the same class.

5. The experimental results showed that the proposed approach WOA-SVM is not time-consuming and provided a high classification accuracy rate compared to other feature selection approaches.

5 Conclusion

In this paper, we propose a novel approach for feature/band selection and parameter determination of support vector machine applied in the context of band selection in hyperspectral image classification and feature selection in 2D image. The proposed approach is called WOA-SVM and it used the whale optimization algorithm which is a recent meta heuristic inspired from the social behaviors of humpback whale. A new objective function based on the SVM weight vector and a new measure has been developed. The paper presents two contributions: the first one is the tuning of SVM parameter C and the Gaussian kernel parameter σ. The second is to select the relevant and informative features/bands which provide the high classification accuracy rate. The proposed approach was applied on three widely used hyperspectral images: Salinas, Indian Pines, and Pavia University. In addition, we propose to conduct a second experimentation under MIAS dataset which is a set of mammogram images. The experimental results show that WOA-SVM produced satisfactory results and was adequate to determine the optimal parameters of SVM. The stability of WOA algorithm was verified. We executed the algorithm 100 times and computed the classification accuracy rate. The results indicate that the proposed approach is very stable.

WOA-SVM was compared with several filter and wrapper approaches, and two classifier using all the features. The classification accuray rate obtained by WOA-SVM indicated that this approach performs significantly advantageously with regards to other. These results provide encouragement to analysis WOA-SVM in other application such as the gene selection in DNA microarray dataset for cancer diagnosis.

Table 5. Classification accuracy rate and computational time obtained by WOA-SVM using two objective function.

Accuracy rate (%) : WOA-SVM	Time				
		$fitness\ J_1$			
$fitness\ J_2$					
	Best	Mean	Worst	Best	Mean
Worst					
MIAS	100	98.82	94.91	100	98.71
96.61	2.81				

Table 6. Comparison of classification accuracy rate obtained by BGWO and other approaches for MIAS dataset.

MIAS : Database - Accuracy (%)	
Filter Approaches	
MIM	67.52
mRMR	92.37
MIFS	76.27
JMI	98.83
CMIM	76.27
Gini Index	86.44
Relief	99.92
Wrapper Approaches	
PSO	100
BGSA	89.32
BBA	91.18
GA	95.96
Using All Features	
KNN	93.62
SVM	99.14
This study	
WOA-SVM J_1	100
WOA-SVM J_2	100

References

1. **Cortes, C., Vapnik, V. (1995).** Support-vector networks. Machine Learning, Vol. 20, No. 3, pp. 273–297.

2. **Dai, J., Liu, Q., Zou, X., Zhang, C. (2024).** Feature selection based on fuzzy combination entropy considering global and local feature correlation. Information Sciences, Vol. 652, pp. 119753. DOI: https://doi.org/10.1016/j.ins.2023.119753.

3. **Dong, J., Li, X., Zhao, Y., Ji, J., Li, S., Chen, H. (2024).** An improved binary dandelion algorithm using sine cosine operator and restart strategy for feature selection. Expert Systems with Applications, Vol. 239, pp. 122390. DOI: https://doi.org/10.1016/j.eswa.2023.122390.

4. **Ethem, A. (2004).** Introduction to machine learning. Second Edition.

5. **Huang, C. L., Wang, C. J. (2006).** A ga-based feature selection and parameters optimization for support vector machines. Expert Systems with Applications, Vol. 31, No. 2, pp. 231–240.

6. **Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., Murthy, K. R. K. (2001).** Improvements to platt's smo algorithm for svm classifier design. Neural Computation, Vol. 13, No. 3, pp. 637–649.

7. **Liu, Y., Wang, G., Chen, H., Dong, H., Zhu, X., Wang, S. (2011).** An improved particle swarm optimization for feature selection. Journal of Bionic Engineering, Vol. 8, No. 2, pp. 191–200.

8. **Matthias, R. (2001).** Algorithms and architectures for machine learning based on regularized neural networks and support vector approaches. Shaker Verlag.

9. **Medjahed, S. A., Fatima, B. (2024).** Applying support vector machines with different kernel to breast cancer diagnosis. Computación y Sistemas, Vol. 28, No. 2, pp. 659Â–667.

10. **Medjahed, S. A., Fatima, B. (2024).** Narwhal optimizer: A novel nature-inspired meta-heuristic algorithm. International Arab Journal of Information Technology, Vol. 21, No. 3, pp. 418–426.

11. **Medjahed, S. A., Fatima, B. (2024).** On the performance assessment and comparison of features selection approaches. Computación y Sistemas, Vol. 28, No. 2, pp. 607Â–622.

12. **Medjahed, S. A., Saadi, T. A., Benyettou, A. (2016).** Microcanonical annealing and threshold accepting for parameter determination and feature selection of support vector machines. Journal of Computing and Information Technology, Vol. 24, No. 4, pp. 369–381.

13. **Medjahed, S. A., Saadi, T. A., Benyettou, A., Ouali, M. (2015).** Binary cuckoo search algorithm for band selection in hyperspectral image classification. IAENG International Journal of Computer Science, Vol. 42, No. 3, pp. 183–191.

14. **Medjahed, S. A., Saadi, T. A., Benyettou, A., Ouali, M. (2015).** Binary cuckoo search algorithm for band selection in hyperspectral image classification. IAENG International Journal of Computer Science, Vol. 42, No. 3, pp. 183–191.

15. **Medjahed, S. A., Saadi, T. A., Benyettou, A., Ouali, M. (2016).** Gray wolf optimizer for hyperspectral band selection. Applied Soft Computing Journal, Vol. 40.

16. **Medjahed, S. A., Saadi, T. A., Benyettou, A., Ouali, M. (2016).** Gray wolf optimizer for hyperspectral band selection. Applied Soft Computing, Vol. 40, pp. 178–186.

17. **Medjahed, S. A., Saadi, T. A., Benyettou, A., Ouali, M. (2016).** A new post-classification and band selection frameworks for hyperspectral image classification. The Egyptian Journal of Remote Sensing and Space Sciences, Vol. 19, No. 2, pp. 163–173.

18. **Medjahed, S. A., Saadi, T. A., Benyettou, A., Ouali, M. (2017).** Kernel-based learning and feature selection analysis for cancer diagnosis. Applied Soft Computing Journal, Vol. 51, pp. 39–48.

19. **Mirjalili, S., Lewis, A. (2016).** The whale optimization algorithm. Advances in Engineering Software, Vol. 95, pp. 51–67.

20. **Mirjalili, S., Mirjalili, S. M., Lewis, A. (2014).** Grey wolf optimizer. Advances in Engineering Software, Vol. 69, No. 1, pp. 46–61.

21. **Mirjalili, S., Mirjalili, S. M., Yang, X.-S. (2013).** Binary bat algorithm. Neural Computing and Applications, Vol. 25, No. 3-4, pp. 663–681.

22. **Platt, J. C. (1999).** Fast training of support vector machines using sequential minimal optimization. Advances in Kernel Methods - Support Vector Learning, pp. 185–208.

23. **Qian, W., Xiong, Y., Ding, W., Huang, J., Vong, C.-M. (2024).** Label correlations-based multi-label feature selection with label enhancement. Engineering Applications of Artificial Intelligence, Vol. 127, pp. 107310. DOI: https://doi.org/10.1016/j.engappai.2023.107310.

24. **Rashedi, E., Nezamabadi, H., Saryazdi, S. (2009).** Gsa: A gravitational search algorithm. Information Sciences, Vol. 179, No. 13, pp. 2232–2248.

25. **Rashwan, S., Dobigeon, N. (2017).** A split-and-merge approach for hyperspectral band selection. IEEE Geoscience and Remote Sensing Letters, Vol. 14, No. 8, pp. 1378–1382.

26. **Rychetsky, M. (2001).** Algorithms and architectures for machine learning based on regularized neural networks and support vector approaches. Shaker Verlag GmBH, Germany.

27. **Sewwandi, M. A. N. D., Li, Y., Zhang, J. (2024).** Granule-specific feature selection for continuous data classification using neighborhood rough sets. Expert Systems with Applications, Vol. 238, pp. 121765. DOI: https://doi.org/10.1016/j.eswa.2023.121765.

28. **Su, H., Cai, Y., Du, Q. (2017).** Firefly-algorithm-inspired framework with band selection and extreme learning machine for hyperspectral image classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 10, No. 1, pp. 309–320.

29. **Sun, K., Geng, X., Ji, L., Lu, Y. (2014).** A new band selection method for hyperspectral image based on data quality. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 7, No. 6, pp. 2697–2703.

30. **Taylor, J. S., Cristianini, N. (2004).** Kernel Methods for Pattern Analysis. Cambridge University Press.

31. **William, F. G., Steve, L. (2002).** Efficient svm regression training with smo. Journal Machine Learning, Vol. 46, No. 3, pp. 271–290.

32. **Yuan, Y., Zhu, G., Wang, Q. (2015).** Hyperspectral band selection by multitask sparsity pursuit. IEEE Transactions on Geoscience and Remote Sensing, Vol. 53, No. 2, pp. 631–644.

Article received on 29/07/2024; accepted on 17/01/2025.
**Corresponding author is Seyyid Ahmed Medjahed.*