Binary Cuckoo Search Algorithm for Band Selection in Hyperspectral Image Classification

Seyyid Ahmed Medjahed, *Member IAENG*, Tamazouzt Ait Saadi, Abdelkader Benyettou, and Mohammed Ouali

***Abstract*— Spectral band selection has been a very active and important field of research in hyperspectral image classification. The hyperspectral images contain redundant measurements and irrelevant information which reduce significantly the classification accuracy rate. In this paper, we propose a new framework for band selection problem based on Binary Cuckoo Search. We cast the problem of band selection as a combinatorial optimization problem and we use a binary version of Cuckoo Search algorithm which is a new metaheuristic algorithm more efficient than practical swarm optimization and genetic algorithms. The experiments were applied on three widely used benchmark hyperspectral data sets. The proposed approach was used under the k-nearest neighbor classifier and compared with several feature selection algorithms defined in the literature. The results show that the proposed approach provides a high classification accuracy rate in comparison to other approaches by using a few samples for training and a small number of bands.**

***Index Terms*— Spectral band selection, hyperspectral image classification, binary cuckoo search, optimization**

I. Introduction

RECENTLY, hyperspectral image classification has been very interesting field of research in many applications [1-3]. It consists to classify each pixel in a specific label. The hyperspectral images are composed of hundreds of bands with very high resolution taken under different frequencies. The National Aeronautics and Space Administration (NASA) Jet Propulsion Laboratory is the first that has used the hyperspectral images and builds the Airborne Visible-Infrared Imaging Spectrometer (AVIRIS). The AVIRIS can records hyperspectral image using more than 200 spectral bands with very high resolution and high-dimensional image data [4-5].

In the classification of hyperspectral images, several problems should be taken on consideration such as: the quality of bands and the high number of spectral channels. However, the application of the band/feature selection approach can significantly decreases the computational time and space dimension. Also, the selection of relevant bands extremely improves the classification accuracy rate. The aims of the band selection are to find the most important information from a given set of band by removing the irrelevant and the highly correlated bands.

Feature selection approaches are classified into two general groups: Filter and Wrapper approaches. The first defined a heuristic scoring to measure the quality of features without using classifiers. The filter approaches are very fast and more practical on high-dimension data but they ignore the feature dependencies. The Wrapper approaches are based on the use of classifiers. These approaches are very slow because they must repeatedly call the classifier algorithm, but, they reach a high classification accuracy rate compared to filter approaches [6-8].

In this paper, we model the band selection problem as a combinatorial optimization problem in which a set of features that leads to the best feature space separability is then employed to map the original data to a new one. To solve the combinatorial problem, we use a new version of Cuckoo Search (CS) algorithm which is a one of recently nature-inspired metaheuristic algorithms developed in 2009 by Xin-She Yang of Cambridge University and Suach Deb of C. V. Raman College of Engineering [9]. Recent studies show that CS algorithm is potentially far more efficient than practical swarm optimization and genetic algorithms [10]. The effectiveness of the proposed approach is demonstrated on three hyperspectral image data sets: Indian Pines Scene, Salinas Scene and Pavia University Scene. We compared the obtained results with several feature selection approaches and several classifiers without feature selection. The remainder of the paper is organized as follows: Section II describes and presents the proposed approach. In Section III, we discuss the experimental results. In Section IV, conclusions are drawn with some perspectives.

II. The Proposed Band Selection Approach

The purpose of the proposed approach is to reformulate the band selection problem as a combinatorial optimization problem defined as follows:

Let $b=\{b_1,b_2,...,b_d\}$ the band set of hyperspectral image dataset D with d features.

Let $x=\{x_1,x_2,...,x_d\}$ a binary vector with $x_i \in \{0,1\}$. We define the pair (b_i,x_i) with:

Manuscript received December 30, 2014; revised March 24, 2015.

Seyyid Ahmed Medjahed is with the Institute of Mathematics and Computer Science, University of Science and Technology Mohamed Boudiaf, Oran 31000, Algeria (corresponding author to provide phone: +213-41-56-0333; fax: +213-41-56-0333; e-mail: seyyid.ahmed@univ-usto.dz).

Tamazouzt Ait Saadi is with Computer Science Department, University of Le Havre, 25 Rue Philippe Lebon, 76600 Le Havre (e-mail: tamazouzt.ait.saadi@univ-lehavre.fr).

Abdelkader Benyettou is with the Institute of Mathematics and Computer Science, University of Science and Technology Mohamed Boudiaf, Oran 31000, Algeria. He is the chair of SIMPA Laboratory (e-mail: aek.benyettou@univ-usto.dz).

Mohammed Ouali is with the Department of Computer Engineering, College of Computers and Information Technology, Taif University, K.S.A. (permanent e-mail @: mohammed.ouali@usherbrooke.ca).

(Revised online publication: 18 August 2015)

- If $x_i = 1$ the band b_i will be selected and used for training,
- Else $x_i = 0$ the band b_i will not be used.

The goal is to find a subset of band that minimizes the objective function $f(x)$ which represents the classification error rate. The error rate can be calculated by used any classifier, such as: Support Vector Machine, K-Nearest Neighbor, Optimum Path Forest, or, by using the spectral distance function (SAM, SID, NCC, etc).

The quality of the results depends largely of the search strategy used to solve the combinatorial problem. In this study, we propose to use a binary version of Cuckoo Search Algorithm.

Cuckoo Search is a novel evolutionary optimization algorithm that has been developed by Xin-She and Deb in 2009. The basic idea of this algorithm is that some species of Cuckoo can lay down their eggs in a host nests, though they may remove others eggs to increase the hatching probability of their own eggs [9]. Quite a number of species engage the obligate brood parasitism by laying their eggs in the nests of other host birds [10]. From this idea, the Cuckoo Search algorithm is developed and can be described using the following three rules:

(1) Each cuckoo lays one egg in a randomly nest.
(2) The best nest with high quality eggs will carry over to the next generations.
(3) The number of available hosts nests is fixed, and the egg laid by a cuckoo is discovered by the host bird with a probability $\rho \in [0,1]$. In this case, the cuckoo egg it can throw the egg away or abandon the nest, and build a completely new nest.

A. The Proposed Binary Cuckoo Search Algorithm for Band Selection

Rodrigues *et al.* [11] proposed a binary version of Cuckoo Search Algorithm namely BCS (Binary Cuckoo Search) and used it for theft detection in power distribution systems. In this study, we adopt the BCS algorithm for band selection problems in hyperspectral image classification. The algorithm selects the relevant and the smallest subset of bands that decreases the classification error rate. This version is defined as follows:

```
Input: The hyperspectral image dataset D, number of nests m, number of bands d, number of iterations T.
Output: The subset of selected bands.
For each nest n_i from 1 to m do
    For each band b_j from 1 to d do
        x(i,j) ← Random{0,1}
    End For
End For
f ← 1
For each iteration t from 1 to T do
    For each nest n_i from 1 to m do
        D ← hyperspectral image data
        For each band b_j from 1 to d do
            If x(i,j) = 0 then
                D' ←D-{the band j}
            End If
        End For
        Split pixels of D' into: training and testing set
        Train the classifier with the training set and evaluate it over testing set
        Calculate the classification error rate f'
        If (f' < f) then
            For each band j from 1 to d do
                x'(i,j)←x(i,j)
            End For
        End If
    End For
    For each nest n_i from 1 to m do
        For each band b_j from 1 to d do
            Select the worst nests according to ρ ∈ [0,1] and replace them for new solution
        End For
    End For
    For each nest n_i from 1 to m do
        For each band b_j from 1 to d do
            step←μ/|υ^(1/β)|
            x(i,j) ← x(i,j) + step
            If (x(i,j) > σ) then
                x(i,j) ← 1
            Else
                x(i,j) ← 0
            End If
        End For
    End For
End For
```

The algorithm starts with a random solution; each nest contains a vector of binary values randomly generated, and the initial value of the objective function is set to *1* (classification error rate = *1*). In each iteration of the algorithm and for each solution in the nest, the algorithm constructs a new hyperspectral image by removing the bands that have $x_i = 0$ (bands which are not selected). A new training and testing sets are constructed and we use a classifier to calculate the classification error rate. In this phase, the algorithm evaluates the objective function and store the best nest. The next step is to select the worst nests by using the probability ρ and replace them with random nests. The final loop updates the nests by generating a solution via the Mantegna's algorithm [12]:

$$step \leftarrow \frac{\mu}{\left| v^{\frac{1}{\beta}} \right|} \tag{1}$$

where μ and υ are generated from normal distribution.

B. The Objective Function

The objective function evaluates candidate subsets and returns a measure of their performance. The objective functions are divided in two classes:

- The first one is used in the filters approaches; it evaluates the subsets of selected features by their information content: distance of classes, information theoretic, statistical dependence, etc.

- The second category is considered for the wrappers approaches. It uses the classifier system to evaluate the selected features.

In this study, we use two objective functions:

First Objective Function: Generally, the classification error rate is used as the principal objective function. The problem of feature selection will be to select the subset of features that minimize the error rate. Is defined as follows:

$$f_1(b) = \left\{ \frac{\text{Incorrectl y Classified Instances}}{\text{Classified Instances}} \right\} \quad (2)$$

The classification error rate is calculated by using a classifier system under the subset of candidate features.

Second Objective Function: We propose to combine two important terms: the balance error rate and the discriminant ability of features. The problem of feature selection consists to select the relevant subset of features which optimize the error rate and well separate the classes in the features:

The balance error rate (BER) is considered as the main criterion in the feature selection problem. It is given by [14]:

$$BER = \frac{1}{2}\left(\frac{\text{positive instances predicted wrong}}{\text{positive instances}} + \frac{\text{negative instances predicted wrong}}{\text{negative instances}}\right)$$

The BER is the average of the errors on each class.

The second term is the F-score measure which calculates the discriminant ability of feature. For a feature i, the F-score measure is defined as follows [14]:

$$F(i) = \frac{\left(\overline{x}_i^{(+)} - \overline{x}_i\right)^2 + \left(\overline{x}_i^{(-)} - \overline{x}_i\right)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}\left(x_{k,i}^{(+)} - \overline{x}_i^{(+)}\right)^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}\left(x_{k,i}^{(-)} - \overline{x}_i^{(-)}\right)^2}$$

where $\overline{x}_i, \overline{x}_i^{(+)}, \overline{x}_i^{(-)}$ are the average of the feature i, positive and negative datasets, respectively; $x_{k,i}^{(+)}, x_{k,i}^{(-)}$ are the feature i of the k positive instance and negative instance respectively [14].

The F-score defined above is considered for the binary features (two classes). To address the problem of multi-classes, we defined the following equation of F-score for a set of selected features:

$$FS(b) = \frac{\sum_{i=1}^{d} b_i \times F(i)}{\sum_{i=1}^{d} F(i)} \quad (3)$$

where b is a binary vector which defined a feature i is selected or not ($b_i=1$ the feature i is selected, $b_i=0$ the feature i is not selected).

The equation (3) calculates the ration of F-score for a set of selected features.

The objective function to optimize is the sum of two terms:

$$f_2(b) = \alpha_{BER} \times BER(b) + \alpha_{FS} \times FS(b) \quad (4)$$

Where α_{BER} and α_{FS} is the weight coefficients of the BER and F-score term respectively.

III. Experimental Results and Discussions

A. Hyperspectral Image Datasets

The performances analysis of the proposed approach was conducted under three real hyperspectral image datasets widely used in the literature [1-5].

The first hyperspectral image is an airborne data set taken from the ROSIS sensor. This scene contains 113 spectral bands in the spectral range from 0.43 to 0.86 µm with 640×340 pixels taken over the University of Pavia in Italy. The Pavia University Scene contains 9 ground truth classes: Asphalt, Meadows, Gravel, Trees, Painted Metal Sheets, Bare Soil, Bitumen, Self-Bloking Bricks, and Shadows.

The second hyperspectral image used in this study was collected by the AVIRIS sensor. This hyperspectral image was taken over the Indian Pines region in Northwestern Indiana which is a mixed of agricultural and forest area. The size of this scene is 145×145 pixels and it is composited of 220 bands in the wavelength range from 0.4 to 2.5 µm. The Indiana Pines Scene consists of 16 ground truth classes, namely: 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.

The last hyperspectral data was acquired over Salinas Valley in Southern California, USA by AVIRIS Sensor. The number of spectral bands is 224 bands over 512 × 217 pixels. The range of spectral is 0.4 to 2.5 µm. The ground truth of this scene differentiates 16 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.

B. Parameters Settings

The parameters of the proposed approach are setting as follows: The number of nests is set to 20. The algorithm stops when the value of objective function is 0, or, when the algorithm reaches the total number of iterations T=100.

The classifier used to calculate the objective function in the Binary Cuckoo Search for band selection is the k-nearest neighbour (KNN) algorithm. The KNN algorithm is used with the Euclidean distance and k=7 (7 nearest neighbours).

For each classifier system, the number of samples used for training and test phases must be determined. In this study, we consider (10%) of pixels for the training and the remaining pixels (90%) are considered for the test and validation. To overcome the problem of overfiting, we propose to use split the 90% of pixels to two sets: test set and validation set. The validation set is used under the final subset of band, and, the test set is used under the BCS algorithm. The table 1 presents the number of pixels considered for training and testing in each hyperspectral images.

TABLE 1. INFORMATION ABOUT THE NUMBER OF PIXELS USED IN TRAINING AND TESTING PHASES FOR EACH HYPERSPECTRAL IMAGE.

Class	Hyperspectral Images		
	Pavia University	Indian Pines	Salinas
Number of Classes	9	16	16
Total Number of Pixels	42776	10249	54129
Training set	4277	1024	5412
Test set	38499	9225	48717

C. Experimental Results

In this section, we present the experimental results obtained by the proposed approach. The experimentations are conducted in terms of: overall accuracy (OA) which is the number of correctly classified pixels with respect to the total number of test pixels, average accuracy (AA) and Individual Class Accuracy (ICA). The stability measures are also used.

We compared the proposed approach with five filter feature selection techniques: mRmR, CMIM, JMI, MIFS, Relief, and with the most widely feature selection method used in the literature: GA (Genetic Algorithm for Feature Selection) which is a wrapper technique. To validate the performances our approach, we compared it with tow classifier systems without using feature selection techniques. We consider: the SVM and KNN classifier systems by using all the features.

The table 2 illustrates the results OA, AA and individual class accuracies obtained under the Pavia University Scene hyperspectral image.

Table 2 presents the classifications accuracy rate provided by the proposed approach and other methods. The columns represent the different methods used in the comparison protocol and the last column contains the results of our approach. The rows represent the classes of the hyperspectral images and the two last columns contain respectively the average accuracy and overall accuracy.

We clearly note that in the challenging classification scheme, the proposed approach reaches better results and good performances that the other methods. Furthermore, the proposed approach significantly improves the OA and AA. Compare with SVM, the results are slightly similar in term of AA.

The BCS algorithm is very performance when is used with the objective function f_2. The good results are recorded for the objective function f_2 with 90,17% of average accuracy and 92,60% of overall accuracy.

TABLE 2. OA (%), AA (%), INDIVIDUAL CLASS ACCURACIES (%) OBTAINED BY THE PROPOSED APPROACH AND COMPARED WITH DIFFERENT FEATURE SELECTION APPROACHES AND TWO CLASSIFIERS WITHOUT FEATURE SELECTION APPLIED TO THE PAVIA UNIVERSITY HYPERSPECTRAL DATA SET BY USING 10% OF SAMPLES AS TRAINING SET.

		Feature Selection Approaches						All the bands		This Study	
#	Class	mRmR	cmim	jmi	mifs	Relief	GA	SVM	KNN	Objective Function f_1	Objective Function f_2
1	Asphalt	83,39	79,89	80,08	81,39	83,52	82,28	84,93	86,53	88,66	90,03
2	Meadows	95,82	93,17	94,74	95,07	94,18	96,91	70,79	97,93	98,44	99,21
3	Gravel	53,21	53,93	56,43	40,06	47,38	43,75	67,16	70,00	74,59	80,00
4	Trees	77,20	60,03	61,38	59,09	53,47	79,73	97,77	83,12	83,57	85,34
5	Painted Metal Sheets	99,54	98,88	98,88	98,88	89,78	99,16	99,46	99,08	98,98	99,55
6	Bare Soil	48,78	26,47	27,16	52,76	62,72	58,25	92,83	58,83	71,03	75,09
7	Bitumen	81,11	78,67	79,70	69,27	75,75	75,47	90,42	85,62	87,69	90,87
8	Self-Bloking Bricks	85,71	81,47	81,47	81,87	79,97	83,47	92,78	87,58	89,21	91,56
9	Shadows	100	75,33	82,45	100	100	100	98,11	99,87	99,87	99,91
	AA	**80,53**	**71,98**	**73,59**	**75,38**	**76,31**	**79,89**	**88,25**	**85,40**	**88,00**	**90,17**
	OA	**83,82**	**77,29**	**78,50**	**80,99**	**81,81**	**84,57**	**81,01**	**87,95**	**90,38**	**92,60**

TABLE 3. OA (%), AA (%), INDIVIDUAL CLASS ACCURACIES (%) OBTAINED BY THE PROPOSED APPROACH AND COMPARED WITH DIFFERENT FEATURE SELECTION APPROACHES AND TWO CLASSIFIERS WITHOUT FEATURE SELECTION APPLIED TO THE INDIAN PINES HYPERSPECTRAL DATA SET BY USING 10% OF SAMPLES AS TRAINING SET.

		Feature Selection Approaches						All the bands		This Study	
#	Class	mRmR	cmim	jmi	mifs	Relief	GA	SVM	KNN	Objective Function f_1	Objective Function f_2
1	Alfalfa	0,00	16,22	51,35	0,00	0,00	0,00	84,44	21,63	32,44	34,00
2	Corn-on till	31,67	55,12	58,62	31,67	54,42	52,93	36,79	52,60	66,93	74,89
3	Corn-min till	18,37	47,89	42,17	18,37	24,10	28,16	40,67	51,21	59,19	66,85
4	Corn	1,58	40,53	40,53	1,58	8,42	14,74	72,31	31,06	38,43	39,62
5	Grass/pasture	70,28	83,46	74,16	70,28	57,11	82,17	80,40	85,28	83,98	85,82
6	Grass/tree	66,95	93,15	96,75	66,95	86,64	94,35	78,93	96,91	97,09	99,54
7	Grass/pasture-mowed	4,35	86,96	43,48	4,35	0,00	17,39	95,38	78,27	73,92	75,11
8	Hay-windrowed	51,44	96,08	98,17	51,44	99,22	96,87	76,11	98,96	98,96	99,98
9	Oats	0,00	6,25	25,00	0,00	0,00	0,00	100	0	6,25	56,70
10	Soybeans-no till	13,24	66,84	68,38	13,24	41,65	59,38	53,80	70,83	80,60	82,90
11	Soybeans-min till	62,07	75,51	74,95	62,07	67,92	74,13	39,73	77,75	81,88	84,33
12	Soybeans-clean till	11,58	45,47	38,11	11,58	12,63	32,00	49,12	44,43	54,74	58,58
13	Wheat	66,46	92,07	94,51	66,46	91,46	94,51	91,42	96,35	95,74	97,87
14	Woods	92,98	94,76	92,29	92,98	90,51	92,98	81,31	94,97	95,56	96,58
15	Bldg-grass-tree-drives	10,68	11,33	19,09	10,68	7,44	11,65	47,05	18,13	25,25	42,89
16	Stone-steel towers	18,67	65,33	72,00	18,67	25,33	73,33	94,11	81,34	84,00	85,27
	AA	**32,52**	**61,06**	**61,85**	**32,52**	**41,68**	**51,54**	**70,10**	**62,49**	**67,19**	**73,81**
	OA	**46,59**	**69,48**	**69,17**	**46,59**	**57,67**	**64,86**	**56,42**	**71,01**	**76,70**	**85,26**

TABLE 4. OA (%), AA (%), INDIVIDUAL CLASS ACCURACIES (%) OBTAINED BY THE PROPOSED APPROACH AND COMPARED WITH DIFFERENT FEATURE SELECTION APPROACHES AND TWO CLASSIFIERS WITHOUT FEATURE SELECTION APPLIED TO THE SALINAS HYPERSPECTRAL DATA SET BY USING 10% OF SAMPLES AS TRAINING SET.

		Feature Selection Approaches						All the bands	This Study	
#	Class	mRmR	cmim	jmi	mifs	Relief	GA	KNN	Objective Function f_1	Objective Function f_2
1	Brocoli_green_weeds_1	93,97	98,69	97,01	94,40	91,98	98,26	98,32	98,26	99,08
2	Brocoli_green_weeds_2	95,84	99,56	99,43	91,95	61,86	99,60	99,60	99,63	100,00
3	Fallow	85,39	98,73	98,48	58,51	94,18	98,04	99,30	99,43	100,00
4	Fallow_rough_plow	99,73	99,73	99,73	99,73	89,25	99,73	99,73	99,73	100,00
5	Fallow_smooth	94,91	96,13	95,19	87,63	97,34	95,01	96,45	96,45	96,81
6	Stubble	99,27	99,81	99,84	88,64	98,39	99,94	99,94	99,91	100,00
7	Celery	95,98	99,44	99,13	86,63	61,03	99,37	99,37	99,34	99,41
8	Grapes_untrained	76,54	82,54	79,96	66,26	77,23	79,34	83,25	84,09	84,52
9	Soil_vinyard_develop	97,76	97,92	97,60	84,95	98,57	99,38	99,27	99,29	99,73
10	Corn_senesced_green_weeds	75,60	91,80	88,37	36,71	78,42	91,61	91,69	92,15	92,49
11	Lettuce_romaine_4wk	60,94	93,33	93,68	17,89	92,05	92,05	94,04	94,39	95,06
12	Lettuce_romaine_5wk	86,64	99,74	99,81	44,42	92,54	98,83	99,87	99,87	100,00
13	Lettuce_romaine_6wk	96,45	98,23	97,54	9,41	96,45	97,82	97,82	98,09	98,23
14	Lettuce_romaine_7wk	92,41	94,39	94,51	33,76	92,64	93,57	94,16	94,74	94,83
15	Vinyard_untrained	57,33	60,00	61,99	42,13	47,52	57,88	63,54	64,88	65,45
16	Vinyard_vertical_trellis	76,42	98,27	97,72	28,98	82,16	97,86	98,06	98,06	98,93
	AA	**86,57**	**94,27**	**93,75**	**60,75**	**84,48**	**93,64**	**94,65**	**94,89**	**95,28**
	OA	**83,79**	**89,55**	**88,87**	**66,23**	**79,28**	**88,57**	**90,35**	**90,76**	**91,04**

The results obtained by the SVM and KNN by using all the bands are identical to our approach by using the objective function f_1 with an advantage in overall accuracy.

The second experiment is conducted in Indian Pines hyperspectral image. The obtained results are described in table 3.

The table 3 shows the OA, AA and individual class accuracies obtained by our approach with regard to other methods. All the experimentations have been made with the same training and test sets. Compare with other methods, our approach provides good results and it can be observed that exceeds the others in term of OA. Also, we show that our feature selection approach provides satisfactory results when the objective function f_2 is used. SVM by using the entire band provides results closer to our approach.

Note that in the both table 2 and 3, the results of SVM are reported from [13].

Finally, we conduct experiments on Salinas Scene hyperspectral image. The table 4 shows the obtained results.

Table 4 presents a comparison protocol between the results obtained by our approach and the other methods. In all considered classes, the classification accuracy rate obtained by the proposed approach is better than the corresponding result for other methods.

The best results are observed when the objective function f_2 is used.

Fig.1. and Fig.2. show the AA and OA obtained by our approach and compared to others methods.

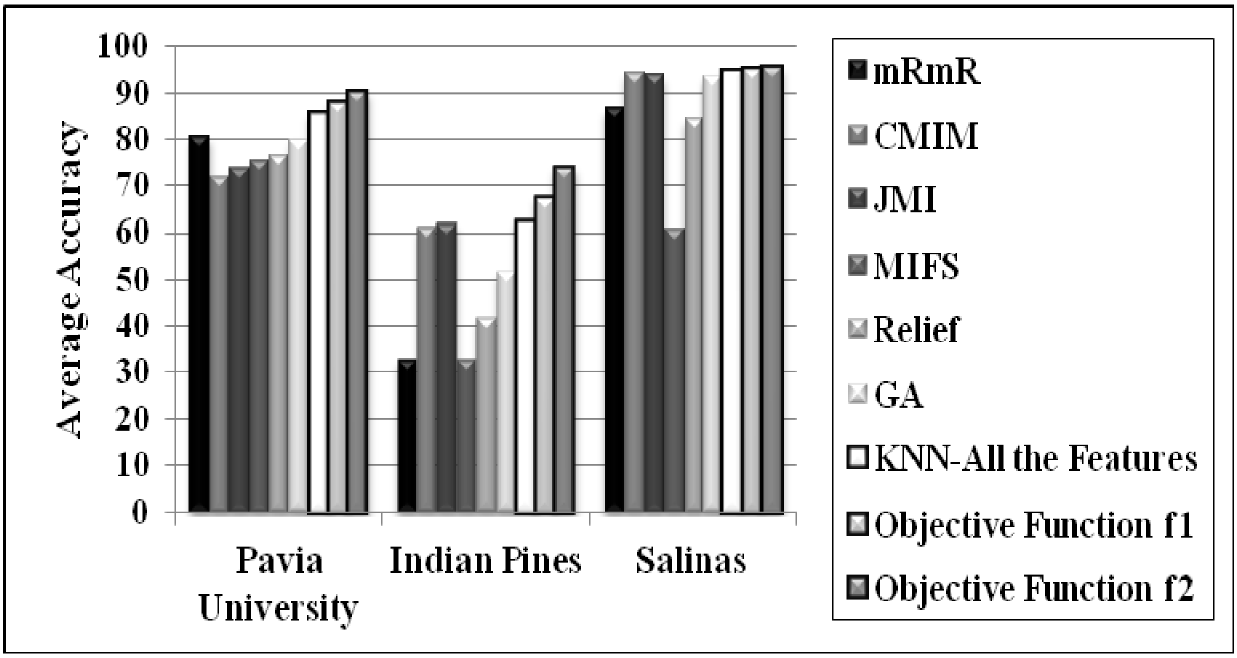

Fig. 1. Average accuracy bars results of Pavia University, Indian Pines and Salinas scene.

(Revised online publication: 18 August 2015)

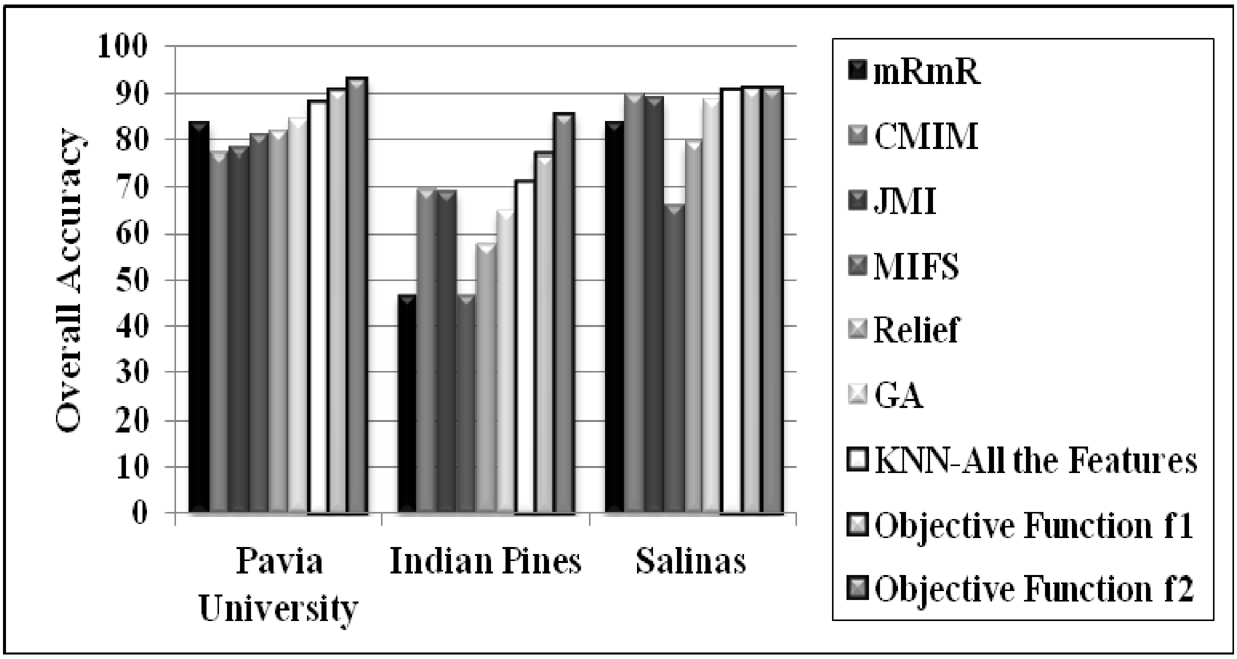

Fig. 2. Overall accuracy bars results of Pavia University, Indian Pines and Salinas scene.

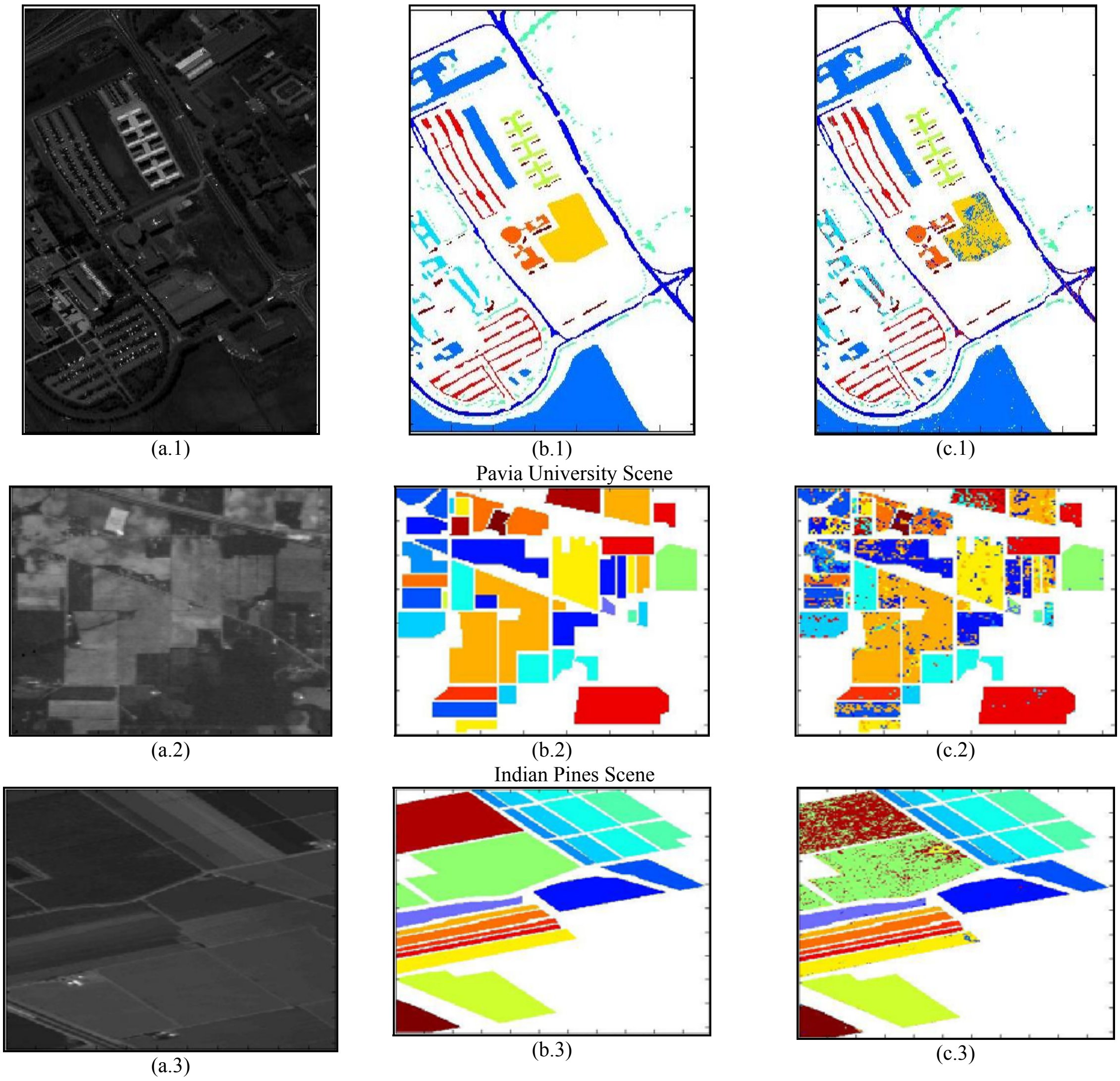

Fig. 3. The classification maps obtained by the proposed approach for each hyperspectral data set by using the objective function f_1.

(Revised online publication: 18 August 2015)

In this study, we have used the hyperspectral pixel classification to evaluate the performance of the proposed band selection approach. All the produced classification results are illustrated in the figures Fig.1. and Fig.2., we analyzed the results by using two different schemas: average accuracy bars and overall accuracy bars. The AA and OA bars represent the comparative results for the Pavia University, Indian Pines and Salinas scenes produced by several feature selection methods.

It is clearly observed from figures Fig.1. and Fig.2. that the results provided by the proposed approach dominates all the other approach. Moreover, the average accuracy and overall accuracy bars of the proposed approach with objective function f_2 and objective function f_1 are the top positions at the rank of classification performance.

As shown in Fig.2. with regard to the Indian Pines scene, the accuracy obtained by the proposed approach are very good and outperform to other methods, with an advantage to the proposed approach used with the objective function f_2. Figure Fig.3. shows the Hyperspectral image visualization results.

The first column of the figure Fig.3. represents the spectral image band number 10 of Pavia University and Indian Pines Scene, and the band number 30 for the Salinas Scene. The second column is the Ground Truth of each hyperspectral image data set. The third column corresponds to the classification map obtained by the proposed approach by using the function f_1.

The visual results are illustrated on the figure Fig.3. and it shows that for both classification maps of Pavia University and the Salinas Scenes are very satisfactory and very similar to the ground truth.

We clearly remark some dissimilarity on Indian Pines classification map compared to it ground truth; this is explained by the small number of training pixels. The Indian Pines hyperspectral image is 145×145 pixels and classified into 16 categories. We have used 10% of ground truth data as training samples which make the learning phase very difficult. For example, the class "Oats" contains 20 pixels, so the number of pixels used for training this class is 2 pixels.

In all hyperspectral maps, the results generated by the proposed approach are encouraged. This is explained by the fact that the proposed approach takes advantage from the binary cuckoo search algorithm that minimizes the classification error rate and attempts to reach the optimum by choosing the relevant bands.

The effectiveness of a feature selection method can be estimated by computing the stability measure. In other terms, the stability is the robustness measure of the feature selection method by selecting different feature subsets produced on different training sets [15]. Several stability measures have been proposed, in this study, we use the S_S stability, S_H stability and Kuncheva stability I_C which are widely used in the literature. These stability measures are defined as follows:

Given two feature subsets s_i and sj, the S_S stability proposed by Kalousis *et al.*[15] is given by:

$$S_S = 1 - \frac{|S_i| + |S_j| - 2|S_i \cap S_j|}{|S_i| + |S_j| - |S_i \cap S_j|} = \frac{|S_i \cap S_j|}{|S_i \cup S_j|} \qquad (5)$$

With | . | is the cardinality.

Dunne *et al.* [16] defined the stability as the relative Hamming distance. The S_H stability is defined as:

$$S_H = 1 - \frac{|S_i \setminus S_j| + |S_j \setminus S_i|}{n} \qquad (6)$$

The value of S_S and S_H are between [0,1].

In [17], Kuncheva define the stability I_C as the consistency index for two feature subsets:

$$I_C = \frac{r - \frac{k^2}{n}}{k - \frac{k^2}{n}} = \frac{rn - k^2}{k(n-k)} \qquad (7)$$

With $k = |S_i| = |S_j|$ and $r = |\, S_i \cap S_j \,|$

The subsets must have the same cardinality; the value of I_C is between 1 and -1.

The stability measure of more than two subsets is the average of all pairwise.

To measure the stability of our proposed approach, we run the BGWO algorithm with the objective function f_2 20 times under 20 different training sets for each hyperspectral image.

The results of stabilities measures are plotted in the following figures.

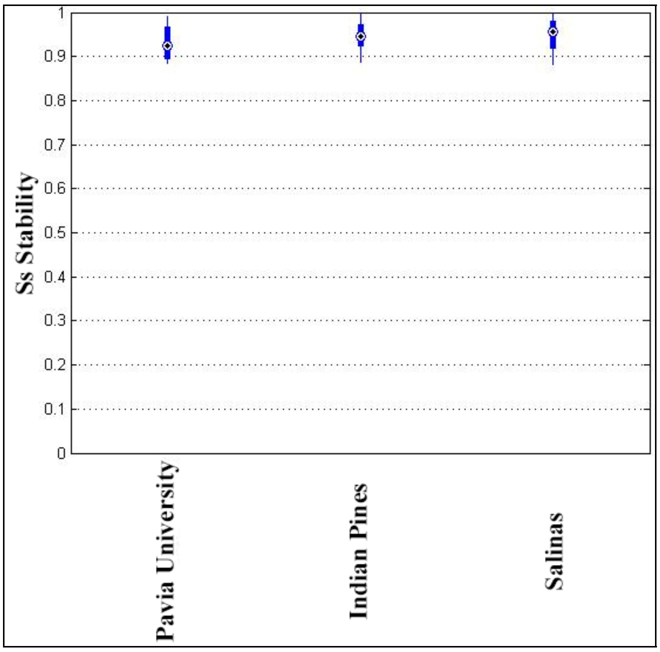

Fig. 4. S_S Stability for Pavia University, Indian Pines and Salinas hyperspectral image.

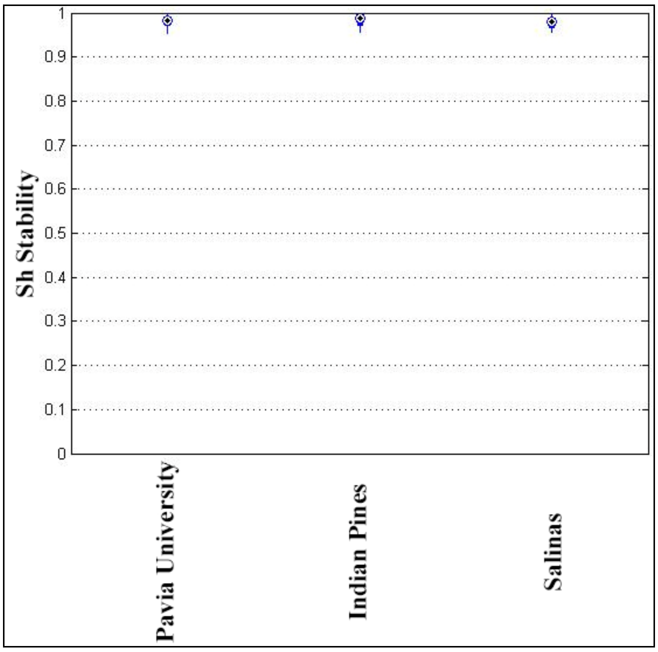

Fig. 5. S_H Stability for Pavia University, Indian Pines and Salinas hyperspectral image.

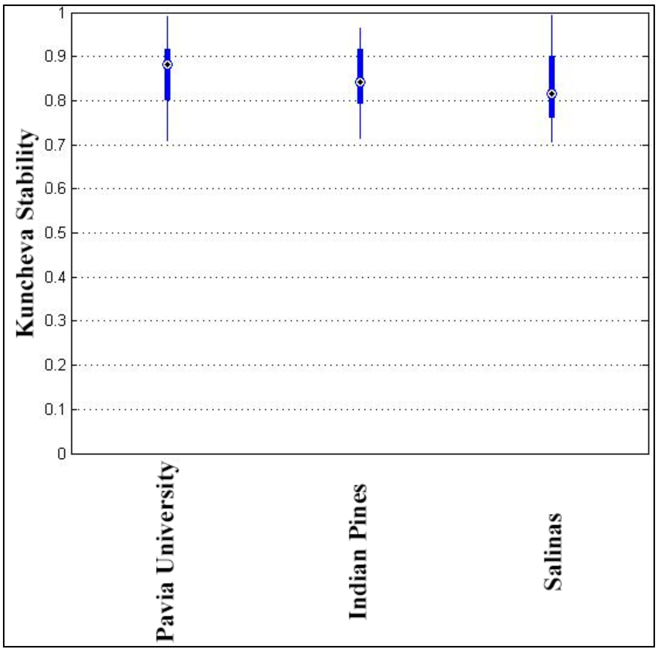

Fig. 6. I_C Stability for Pavia University, Indian Pines and Salinas hyperspectral image.

Figures 4, 5 and 6 show the S_S, S_H and Kuncheva stability obtained by our approach for each hyperspectral image. The blue box indicates the upper and the lower quartiles. The small circle represents the median values and the blue line indicates the maximum and the minimum values. The stabilities obtained for each hyperspectral image have values between 0.9 and 1. A value close to 1 means that our proposed approach is very stable and remains unchanged against the different training sets. Also, the proposed approach produces almost the same features subset for different training sets.

Figures 7, 8 and 9 illustrate the Stability versus the mean overall accuracy.

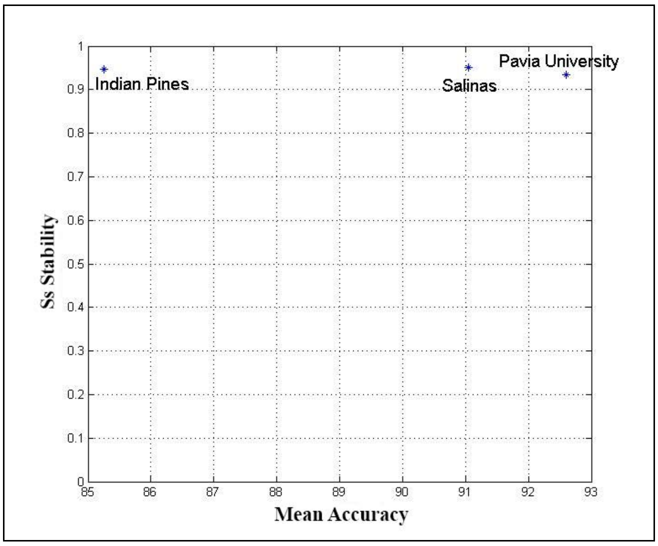

Fig. 7. S_S Stability versus the mean overall accuracy for Pavia University, Indian Pines and Salinas hyperspectral image.

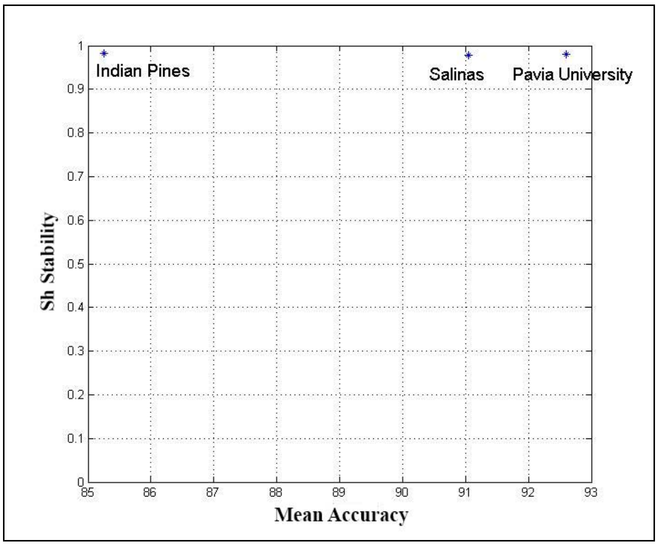

Fig. 8. S_H Stability versus the mean overall accuracy for Pavia University, Indian Pines and Salinas hyperspectral image.

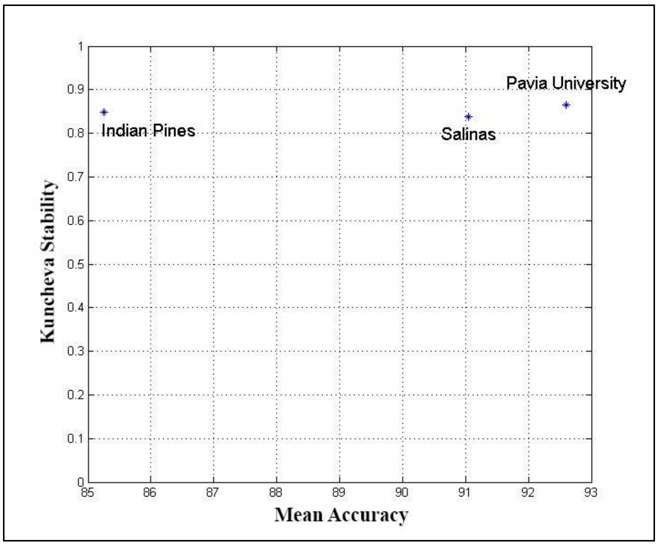

Fig. 9. I_C Stability versus the mean overall accuracy for Pavia University, Indian Pines and Salinas hyperspectral image.

Figures 6, 7 and 8 show the S_S, S_H and Kuncheva stability obtained by proposed approach versus the mean overall accuracy for each hyperspectral image.

A robust feature selection approach is an approach that has stability close to 1 and produces a high classification accuracy rate. We clearly show in the figures 6, 7 and 8 that the results obtained by our approach appear in the top of the figures.

Finally, we can say that the proposed approach is very stable and provides the same subsets of features regardless of the training set and produces a high classification accuracy rate.

VI. Conclusion

In this paper, a new framework for band selection in hyperspectral image classification has been proposed. The proposed approach is based on the new optimization approach called binary cuckoo search algorithm which is a binary version of cuckoo search algorithm. The problem of band selection has been reformulated as combinatorial problem and the objective function to minimize is the classification error rate. The binary cuckoo search is used to optimize the problem. Very recent studies have shown that the cuckoo search is very efficient than practical swarm optimization and genetic algorithms. Experiments have been carried out on three real hyperspectral image data sets: Pavia University, Indian Pines and Salinas Scene. For the three considered data sets, the analysis of the results shows that the proposed approach provides satisfactory results with regard to other feature selection approaches and classifier systems that used all the feature space.

The proposed approach is more robust and reliable than the other methods when the objective function f_2 is used. Although the results obtained by our approach are very good. We also conclude that this approach is suitable for band selection problem with small number of training instances.

Acknowledgment

Dr. Mohammed Ouali was supported by a Taif University grant (grant number 1-436-4596).

References

[1] A.P. Marco, C. Pablo and R. Miguel, "The effect of spectral and spatial degradation of hyperspectral imagery for the Sclerophyll tree species classification," *International Journal of Remote Sensing*, vol. 34, no. 20, 2013, pp. 7113-7130.

[2] X. Zhiyong, "A spectral–spatial similarity-based method and its application to hyperspectral image classification," *Remote Sensing Letters*, vol. 5, no. 2, 2014, pp.122-130.

[3] B. H. Dora, A. Francisco and Q. B.Pablo, "Exploring ELM-based spatial–spectral classification of hyperspectral images," *International Journal of Remote Sensing*, vol. 32, no., 2, 2014, pp. 401-423.

[4] L. Yang, S. Yang, P. Jin and R. Zhang, "Semi-Supervised Hyperspectral Image Classification Using Spatio-Spectral Laplacian Support Vector Machine," *IEEE on Geoscience and Remote Sensing Letters*, vol. 11, no. 3, 2014, pp. 651-655.

[5] M. A. Hossain, X. Jia and M. Pickering M, "Subspace Detection Using a Mutual Information Measure for Hyperspectral Image Classification," *IEEE on Geoscience and Remote Sensing Letters*, vol. 11, no. 2, 2014, pp. 424-428.

[6] P. Lin, N. Thapa, I. Omer, L. L. and J. Zhang, "Feature Selection: A Preprocess for Data Perturbation*", IAENG International Journal of Computer Science*, vol. 38, no. 2, pp. 168-175, 2011.

[7] A. S. Muthanantha and S. Ramakrishnan, "An Optimized Extreme Learning Machine for Epileptic Seizure Detection*", IAENG International Journal of Computer Science*, vol. 41, no. 4, pp. 212-221, 2014.

[8] C. J. Tu, L. Chuang, J. Y. Chang, and C. H. Yang, "Feature Selection using PSO-SVM", IAENG *International Journal of Computer Science*, vol. 33, no. 1, pp. 111-116, 2007.

[9] Y. Xin-She and S. Deb, "Cuckoo Search via Lévy flights," World Congress on Nature & Biologically Inspired Computing, 2009, pp.210-214. doi:10.1109/NABIC.2009.5393690.

[10] Y. Xin-She,"Cuckoo Search and Firefly Algorithm Theory and Applications," 2014, Springer.

[11] D. Rodrigues, L. A. M. Pereira, T. N. S. Almeida, J. P. Papa, A. N. Souza, C. C. O. Ramos and Y. Xin-She, "BCS: A Binary Cuckoo Search algorithm for feature selection," BCS: A Binary Cuckoo Search algorithm for feature selection, *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2013, pp. 465-468, doi:10.1109/ISCAS.2013.6571881.

[12] R. N. Mantegna, "Fast Accurate Algorithm for Numerical Simulation of Levy Stable Stochastic Processes," *Physical Review E*, vol.49, no. 5, pp. 4677-4683, 1994.

[13] V. Alberto, B. Jon Atli, C. Jocelyn and J. Christian, "Hyperspectral Image Classification with Independent Component Disciminant Analysis," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 49, no. 12, 2011, pp. 4865-4876.

[14] C. Yi-Wei and L. Chih-Jen, "Combining SVMs with Various Feature Selection Strategies," *Feature Extraction Studies Fuzziness and Soft Computing*, vol. 207, pp. 315-324, 2006.

[15] Kalousis A., Prados J., and Hilario M., "Stability of feature selection algorithms," *In Proc. 5th IEEE International Conference on Data Mining (ICDM'05)*, pp. 218–225, 2005.

[16] Dunne K., Cunningham P., and Azuaje F., "Solution to instability problems with sequential wrapper-based approaches to feature selection," *Technical Report TCD-CS-2002-28, Department of Computer Science, Trinity College, Dublin, Ireland*, 2002.

[17] Kuncheva L. I., "A stability index for feature selection," *In IASTED International Multi-Conference, Artificial Intelligence and Applications*, pp. 390–395, 2007.

Seyyid Ahmed Medjahed received his M.Sc. in Data and Knowledge Engineering from Universite d'Oran, Algeria, in 2011. Currently, he is a PhD candidate at the college of Mathematics and Computer Science, Universite des Sciences et de la Technologie Mohamed Boudiaf USTO-MB, Oran, Algeria. He has also been a lecturer at Universite de Relizane since 2012. His research interests include Machine Learning and Image Processing.

Tamazouzt Ait Saadi is a lecturer and research assistant at Universite de Mostaganem, Algeria since September 2003. She received her engineering degree in computer science in 1989, and M.Sc. in 2003. Currently, she is a PhD candidate at Universite du Havre, France.

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

Mohammed Ouali received the Doctor degree in Real Time Informatics, Robotics, and Automatic Control from Mines ParisTech, France, and the Ph.D. in Mathematics and Computer Science from the University of Sherbrooke, Canada, in 1999. Before joining the academia, Dr. Ouali has spent more than 15 years in the industry as a senior member of technical and scientific staff. Currently, he holds a visiting associate professor position with the University of Taif. His research interests include signal and image processing, computer vision, pattern recognition, and big data analysis.

1. Date of modification: August 18, 2015

2. Brief description of changes: last author affiliation changed to the most current one, an acknowledgment section added, biographies corrected.

(Revised online publication: 18 August 2015)