Incorporated Decision-maker-based Multiobjective Band Selection for Pixel Classification of Hyperspectral Images

Diego SAQUI[1,2], José Hiroki SAITO[1,3], Daniel Caio De LIMA[1], Luis Mariano Del Val CURA[3], Steve Tsham Mpinda ATAKY[4]
[1]*Computer Department, UFSCar – Federal University of São Carlos, São Carlos, Brazil*
[2]*IFMS – Federal Institute of Mato Grosso do Sul, Corumbá, Brazil*
[3]*UNIFACCAMP – University Center of Campo Limpo Paulista, Campo Limpo Paulista, Brazil*
[4]*ÉTS – École de Technologie Supérieure, Université du Québec, Montréal (Québec), Canada*
diego.saqui@ufscar.br

***Abstract*—Hyperspectral images (HIs) are characterized by a higher spectral resolution than other images and have applications in various fields, to wit, medicine, agriculture, mining, among others. Segmentation can be obtained from the pixel classification and it is a powerful tool for object identification. Notwithstanding, the problems of the curse of dimensionality and the demand for computational resources occur due to the number of bands. Techniques that reduce dimensionality, such as genetic algorithms, are promising, but they cannot assure a balance between conflicting objectives such as improving classification and reducing the number of bands. Multiobjective band selection can be applied to search for tradeoff solutions that have this balance. Therefore, in this manuscript, we propose a novel method called Incorporated Decision-Marker-based multiobjective band selection (IDMMoBS) that tries to find tradeoff solutions using spectral and spatial information. In the experiments, the IDMMoBS reduced the number of bands between 85.4% and 85.8% of the total and it outperformed the majority of other methods compared in this criterion. For the pixel classification, the IDMMoBS presented better results than all compared cases taking into account all evaluated metrics using SVM classifier. Accordingly, the IDMMoBS is suitable for band selection.**

***Index Terms*—remote sensing, hyperspectral imaging, image segmentation, image classification, evolutionary computation.**

I. Introduction

Hyperspectral images (HIs) are concerned with the measurement, analysis, and interpretation of spectra acquired from a given scene (or specific object) at a short, medium or long distance by an airborne or satellite sensor [1]. These images are characterized by their spatial and spectral resolutions (number of spectral bands) and contain information of materials beyond the visible spectrum range [2]. The advantage of using HIs is due to their unique bands that facilitate the discrimination of different materials [2-3]. Each pixel has a signature consisting of a large number of bands representing the wavelength and different reflectance values. This spectral signature varies according to the type of the object, allowing tasks of identification and analysis. HIs are useful in many applications, to wit, medical imaging, agriculture, among others [4].

Land cover classification is one of the different tasks in Remote Sensing (RS), modeled as part of a supervised process based on previous Ground Truth (GT), which is concerned with the identification of different land cover objects [1]. This identification is possible through the pixel classification and segmentation of HIs. The segmentation is the process of subdivision of the image into its constituent regions. For pixel classification and subsequent segmentation, algorithms with peculiar statistical properties such as Support Vector Machines (SVM), Gaussian Maximum Likelihood Classifier (GMLC) and Random Forest (RF) are required [5-7]. These algorithms can be used directly for segmentation or can be complemented using image processing techniques.

HIs have a high dimensionality of data due to the number of bands [8], and because of this some challenges are common, such as:

i) Hughes's phenomenon/curse of dimensionality, described by the existence of high dimensionality for a low number of samples. This problem is common in HIs because of the cost and difficulty of elaborating GTs, and it can result in poor classification accuracy [9];

ii) information redundancy caused by the high correlation between neighboring bands;

iii) computation complexity for classifiers based on conventional statistics [10];

iv) the high cost of HIs sensors [11]; and

v) difficulty in data transmission and storage. Modern data capturing devices, such as Drones, cannot have many computational resources to meet the demand for HIs processing.

Thereof, HIs classification and segmentation are often preceded by dimensionality reduction techniques which are categorized as feature selection or feature extraction [12].

Feature extraction performs a feature space transformation for a new space with a smaller dimension, and some examples of this category are principal component analysis (PCA) [13] and independent component analysis (ICA) [14]. Feature selection is a routine of selecting a subset of the original features, resulting in a smaller number of features where the original representation of the variables is not changed. Therefore, feature selection is preferred when the original meaning of the features is demanded [15].

For HIs, there is an interest in maintaining the original

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Digital Object Identifier 10.4316/AECE.2019.04003

meaning of the information of bands. Thus, feature selection is often applied and for this purpose, it is called band selection. The advantage of this routine is the possibility of identifying bands that characterize objects more efficiently than feature extraction, in this light, avoiding problems like the curse of dimensionality. With fewer identified bands, it is also possible to construct specific and cheaper sensors, which require less computational resources for classification, transmission, and storage of obtained data.

There is a great diversity of methods for band selection. Some examples are a clustering-based method developed using each channel map as a data point [16], sparse representation methods used to select the band subset [17-18], and structure-aware measures for band informativeness and independence [19]. A traditional clustering-based method for band selection is Ward's Linkage Strategy using Mutual Information (WaLuMi). The WaLuMi groups the bands according to the mutual information and then it selects a representative band from each group. This was one of the first clustering-based algorithms in the literature and it has shown good results even when compared to recent methods [20]. In [21], the Maximum information and minimum redundancy with clonal selection algorithm (MIMR-CSA) defines a criterion that maximizes the amount of information for the selected band combination while removing redundant information. The authors compare the classification capacity, and the proposed method presented better results than others in the literature.

Methods based on evolutionary algorithms have shown satisfactory results for hyperspectral band selection. Some examples are the Firefly algorithms [22-23], Genetic Algorithms (GAs) [24-25] and Particle Swarm [26-27]. A hybrid method of Wrapper and Filter approaches called Information Gain–Grey Wolf Optimization (IG-GWO) has been proposed, where Information Gain and Gray Wolf Optimizer were combined. IG is an important metric that can measure how much information the features could contribute to the classification. This method was compared with five other state-of-the-art methods and it showed better accuracy results [26]. All these optimization algorithms establish a single objective function, and for band selection application, some measure of segmentation is conducted individually or combined with the reduction of the number of bands in a single objective function. Particularly in [25], a supervised/Wrapper method for band selection based on GA with Support-Vector Machines (GA-SVM) was proposed combining two metrics in fitness function. The GA-SVM uses a weighted sum strategy between accuracy and the inverse of the number of bands to try to find solutions with a balance between these metrics and shows good overall accuracy results. However, the difficulty of this method and the others is to find balanced tradeoff solutions, for example, solutions that simultaneously have a satisfactory classification performance with a reduced number of bands.

The need for reducing the number of bands and improve classification performance makes the band selection problem be characterized as a multiobjective problem since both objectives can be conflicting. For this, multiobjective optimization band selection (MOBS) can be applied [4].

MOBS are usually based on evolutionary algorithms and use the Pareto Frontier that can contain more than one optimal solution. Pareto Frontier (tradeoff) solutions cannot be considered better among each other and therefore cannot be sorted, but these solutions are better than others in the search space.

Different unsupervised MOBS, which can also be categorized as filter methods, can be found in the literature. Non-dominated Sorting Genetic Algorithm 2 (NSGA2) was proposed to find solutions that optimize cumulative texture and spectral information of a band combination using a minimum number of bands [12]. Another method is the Tchebycheff Decomposition Strategy (TDS) for the multiobjective evolutionary algorithm based on decomposition (MOEA/D). This method was studied to optimize the number of bands and entropy contained in the selected band subsets [4].

A state-of-the-art method called incorporated rank-based multiobjective band selection (IRMoBS) using TDS was proposed to optimize three objectives, band number, variance and information entropy [29]. Three band selection methods [17-19], are compared with IRMoBS which presented better classification using SVM, extreme learning machine (ELM) or K-Nearest Neighborhood (KNN) in the pixels of HIs with selected bands.

If the GT is available, supervised/Wrapper methods can be used to fit a classification model based on a priori knowledge; thereby, it is expected to obtain better results in band selection. Wrapper methods usually require a high computational cost but they present better results than filter methods because they incorporate the classifier into the selection process [30]. This type of method, together with multiobjective algorithms, can be an alternative for the band selection problem. Nonetheless, there were little studied in the literature.

Another feature little explored in the literature is the use of spatial information with band selection algorithms. A study of the application of filters for spatial information is on HIs is performed in [31]. In this study the authors show that even simple filtering can greatly improve the performance of classifiers in HIs. Therefore, spatial filtering can be introduced to improve the classification accuracy of HIs.

In this paper, a novel method based on MOEA/D is modeled for MOBS of HIs. The method is called incorporated decision-maker-based multiobjective band selection (IDMMoBS) and aims to deal with the dimensionality of HIs through the search for tradeoff solutions. The IDMMoBS is modeled as a Wrapper strategy and therefore the SVM classifier is used. The IDMMoBS incorporates a decision-maker (DM) to select only one solution from Pareto Frontier and a repair method based on WaLuMi [20] and IG/IG-GWO [26]. Another differential of this study is that the use of spatial information was considered. The DM, the repair technique and the use of spatial information are little explored in band selection, and this is quite innovative in this study.

The main contributions of this manuscript are:

i) A new Wrapper MOBS to search for tradeoff solutions since these types of methods have not been explored enough;

ii) Use of DM and a repair method that are rarely

discussed in band selection methods;

iii) Use of pixel neighborhood information in each band;

iv) Comparisons with other supervised (GA-SVM [25]), semi-supervised (IG-GWO [26]) and unsupervised (WaLuMi [20], MIMR-CSA [21], IRMoBS [29]) state-of-the-art methods;

v) Comparisons with HIs with all bands;

vi) Analysis of the classification of the pixels of HIs with bands selected by the different methods compared, being this classification performed by SVM;

vii) Use of spatial information.

The remainder of this text is organized as follows. Section II presents a detailed description of the proposed method. The experiments are conducted and the results reported in Section III. At last, in Section IV, the main contribution of this manuscript is summarized and future work direction is pointed.

II. METHODOLOGY

Multiobjective optimization is related to solutions with contradictory objectives to be optimized simultaneously searching for tradeoff solutions. Methods for multiobjective optimization are based on the Pareto Frontier and use the non-dominance relation to compare solutions [4][12][29].

A solution x dominates a solution y, if x is better than y in at least one objective and equal or better in all other objectives. The notation x that dominates y is given by $x \succ y$. Solutions that are not dominated by any other are Pareto-optimal solutions and constitute the Pareto Frontier.

A Pareto-optimal solution to a multiobjective optimization problem could be an optimal solution of a scalar optimization problem where the objective is an aggregation of all the objectives to be optimized. Therefore, the approximation of the Pareto Frontier can be decomposed into a number of scalar objective optimization subproblems [32].

There are several approaches for converting the Pareto Frontier approximation problem into scalar optimization subproblems; and one explored in unsupervised MOBS is the TDS [4, 29]. In this paper, the TDS is used in a supervised and Wrapper MOBS to compose the IDMMoBS.

In IDMMoBS proposed herein, each band of a solution generated from a HI is represented by a binary code in a vector as shown in Fig. 1. The value 1 indicates the presence of the band equivalent to that position and 0 otherwise.

0	1	0	0	1	1	0	...	number of bands

Figure 1. Representation of the solution.

Two objective functions, f_1 and f_2, that should be maximized, are designed for IDMMoBS. f_1 is based on the overall accuracy of each solution x and it is shown in eq. (1).

$$f_1(p,\hat{p}) = \frac{1}{n}\sum_{i=1}^{n}(1*(p_i = \hat{p}_i)) \quad (1)$$

where n is the number of pixels, i represents each pixel, p_i is the original value of the class in the GT, and $\hat{p}_i$ is the predict value of the class. The choice for overall accuracy is because it is the most used metric in the studies found to evaluate the pixel classification in images with selected bands and it is directly related to the success rate of classification of pixels.

For the reduction of bands, it is considered in f_2 the difference of 1 and the ratio between the number of bands of solution x and the total number of bands possible, as shown in eq. (2).

$$f_2(x) = 1 - \frac{t(x)}{len(x)} \quad (2)$$

where $len(x)$ is the size of the array of the solution x and $t(x)$ is the total of bands present in the solution x as shown in eq. (3).

$$t(x) = \sum_{n=0}^{len(x)} x_n \quad (3)$$

The TDS used in the IDMMoBS aggregates the m objectives, which in this case are f_1 and f_2, into a single scalar objective function. This aggregation is performed through the difference in fitness values of f_1 and f_2 between a solution x and an ideal point $z^* = [z_1^*,...,z_m^*]^T$. For maximizing values, $z_j^* = \max\{f_j(x) \mid x \in \Omega\}$ for j-th objective and Ω is the decision space.

The single scalar objective function $g^{tds}(x \mid w, z^*)$ shown in eq. (4).

$$g^{tds}(x \mid w, z^*) = \max_{1 \le j \le m}\{w_j \mid f_j(x) - z_j^* \mid\} \quad x \in \Omega \quad (4)$$

where $w = [w_1,...,w_m]^T$ is the weight vector for the m objectives with $w_j \ge 0$ and $\sum_{j=1}^{2} w_j = 1$. $x^* = [x_1^*,...,x_m^*]^T$ is the ideal solution corresponding to the ideal point z^*.

In the subproblem evolution process, the single scalar objective function in eq. (4) of the neighboring solutions and new solutions generated by mutation and crossover operations are calculated to further determine which solutions could be updated. Solutions that have larger g^{tds} values are more likely to be replaced by others that are closer to the ideal solution [29].

A. IDMMoBS

Prior to the execution of the IDMMoBS algorithm, a mean filter was applied to each spectral band to smooth out band noise and also allow the use of pixel spatial information. A mode filter was applied concurrently to the GT with classes of pixels used for training, where the same mask was used in both cases. To apply these filters, a 2x2 mask was used that worked in accordance with the sample capture strategy presented in the experiments section.

The IDMMoBS is shown in Algorithm 1, where in **Step 1 - Initialization**, the population P is initialized from $OrigHI$ (**Step 1.1**) and fitness is established. In this study the WaLuMi algorithm (based on Mutual Information) was used to generate groups and then randomly a band was selected from each group to generate solutions for the initial population P. This procedure decreases the correlation

between bands and it is repeated until a number N of initial solutions are reached.

In **Step 1.2** the EP is set to empty and in the **Step 1.3** the weight vectors W are established. In **Step 1.3** are also computed the Euclidean distances between the weight vectors and then it is assign the NS nearest solutions to each solution i based on the weight vectors creating the neighborhood $NG(i)$. In **Step 1.4**, the ideal point z^* is determined from the maximum fitness values f_1 and f_2 present in the initial population.

In **Step 2,** the update is performed by a $NGEN$ number of generations. In each generation the DM selects a solution from the EP for display. For each subproblem from $i=1$ to N, in **Step 2.1 - Reproduction and Repair**, the neighborhood solutions $NG(i)$ are selected to generate new solutions NI through crossover and mutation operators. After the crossover operation and before the mutation operation, a repair method is applied. The heuristic of the repair method is explained in the following sections. Then the fitness values of the new solutions NI are established.

The ideal point z^* is updated after comparing the objectives of the new solutions NI with the z^* (**Step 2.2**). Then in **Step 2.3**, the objective values are computed based on eq. (4) for the neighborhood solutions $NG(i)$. Among the neighborhood solutions, the ones whose objective values are larger than g^{tds}, that is, those that are more distant from the z^*, are replaced by solution $y' \in NI$. Then, in **Step 2.4**, the EP is updated with non-dominated solutions.

Finally, in **Step 3** the DM selects the solution according to the established heuristic from EP and returns a trained classification model.

Algorithm 1 IDMMoBS

Input:

$NGEN$ (number of generations),

N (population size),

NS (neighborhood size),

GT,

p (threshold of the amount of bands),

$OrigHI$ (Pixels/samples with all bands),

mr (mutation rate)

and cr (crossover rate).

Output:

Solution selected by the DM and the trained classification model.

Step 1 – Initialization:

Step 1.1 – Creates an initial population P generating solutions from $OrigHI$ using strategy with WaLuMi.

* Establish fitness values using (1) and the mean obtained from cross-validation and calculate the fitness for the number of bands using (2) of solutions in P.

Step 1.2 – Creates an external population $EP = \varnothing$.

Step 1.3 – Set of weight vector $W = [w_1, ..., w_N]^T$.

For each $i = 1...N$,

$NG(i) = \{i_1, ..., i_{NS}\}$,

where $w_{i1}, ...w_{iNS}$ are the NS closest weight vectors to w_i.

Step 1.4 – Initialize z^*.

Step 2 – Update:

For $ngen = 1, ..., NGEN$ do (Stopping Criteria),

DM selects and output a solution of EP,

For $i = 1, ..., N$, do,

Step 2.1 – Reproduction and repair:

For each i_a, i_b from $NG(i)$:

Applies the crossover operator and generates new solutions y'_a and y'_b in NI.

Repairs the solutions in NI.

Applies the mutation operator.

* Establish fitness values of NI.

Step 2.2 – Update of z^*:

For each $y' \in NI$,

For each $j = 1, ..., m(objectives)$,

if $f_j(y') > z_j^*$, then

set $z_j^* = f_j(y')$

Step 2.3 – Update of Neighboring Solutions:

For each index $k \in NG(i)$,

if $g^{tds}(y' \mid w_k, z^*) \le g^{tds}(x_k \mid w_k, z^*)$ then

set $x_k = y'$, $f_1(x_k) = f_1(y')$

and $f_2(x_k) = f_2(y')$.

Step 2.4 – Update of EP:

Remove all solutions dominated by y' from EP.

Add y' to EP if no solutions in EP dominate y'.

Step 3 – Ending:

DM selects and outputs a solution of EP and the trained classification model.

B. Support Vector Machines (SVM)

SVM is based on the theory of statistical learning [34] with principles for finding an optimal hyperplane as a decision function in a large space [35][36]. In the case of classification problems with two distinct classes that can be linearly separable, given infinite possibilities for obtaining the separator function; the SVM selects the best one. This is the simplest scenario for classification by the SVM. The SVM considers the best separator, the one that minimizes the generalization error. Thus, the selected separator is the one with the largest margin between two classes. For nonlinear surfaces, the feature vector is mapped to a larger Euclidean feature space using a kernel function in the nonlinear SVM [37].

The SVM algorithm is used to classify the pixels of each solution for later the average accuracy in eq. (1) be applied. During the execution of IDMMoBS, whenever some pixel classification is performed, 3-Fold cross-validation is applied. The pixels are alternately selected from the HI using the bands of the solution being classified and inserted into one of the folds (systematic sampling). Thus, it avoids that the pixels of the same class be allocated only in the same fold and also allows that the pixels of the same class of different regions in a HI be present in the different folds.

SVM was chosen because it is one of the most commonly used algorithms for HIs and does not require assumption about data distribution (nonparametric). One of its disadvantages for use in conjunction with Wrapper strategy is its training time, but this cannot be a serious problem for the band selection process because it is performed prior to the use of bands in practical applications. Its main advantages are the high classification rate, and that this algorithm requires few training samples to operate which is important for HIs applications [38].

C. Repair strategy and Centroid-based DM

In [30] it is discussed that the repair method is to make the solutions feasible. For the band selection problem a maximum band threshold may be desired and therefore for this problem a set of feasible solutions may be desired. Therefore, in this study the following heuristic is used:

i) Based on [20], this strategy applies WaLuMi to group the highly correlated bands using Mutual Information of each solution;

ii) based on [26], selected the band with the highest information gain (IG) of each group;

iii) sort the set of bands according to the IG;

iv) and select the p bands with the highest IG.

This heuristic limits the search for solutions with balanced objectives, but only when they have less than p bands. p acts as a threshold amount of bands. To avoid the high computational cost, the IG of each band can be calculated before the band selection process and the mutual information between the bands has already been calculated to initialize the population in the WaLuMi.

Even with the p threshold, many solutions can emerge at the Pareto Frontier and so a centroid-based decision-maker (DM) is designed. This DM is based in the method proposed in [33] that is based on centroid and a threshold, and it is adapted to the IDMMoBS.

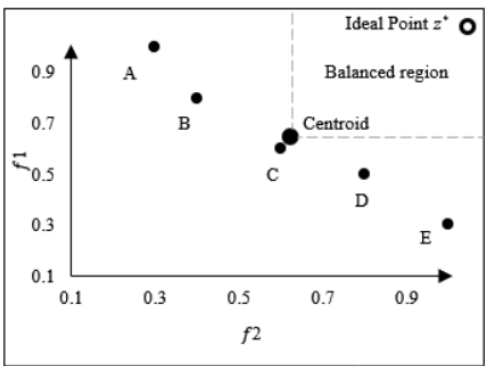


Figure 2. Representation of the DM.

In [33], the centroid is calculated from the f_1 and f_2 of the Pareto-optimal solutions and acts as a threshold to create a region of balanced solutions (balanced region). Fig. 2 shows an example of balanced region (in the dashed region) where the centroid is the largest circle in the lower left corner of the balanced region and the ideal point z^* is the circle without filling. A-E are solutions of the decision space that constitute the Pareto Frontier.

Having the centroid calculated different rules can be established for decision making which solution to choose. In this paper we use the following:

1) If there are solutions in the balanced region then the solution in that region closest to the ideal point z^* is selected;

2) If there is no solution in the balanced region, then the solution in the Pareto Front with value greater than the centroid of one of the functions that presents the best value for another solution is selected.

3) otherwise, the solution of all Pareto-optimal solutions closest to the centroid is selected.

Through this strategy it is possible to select only one solution of the Pareto Frontier.

III. Experiments and Results

A. Experimental setup

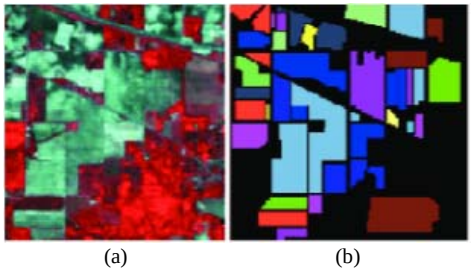
Three different HIs/datasets were used to evaluate the IDMMoBS and compare with other methods. The images used were Indian Pines, Salinas, and Pavia University and information about them are shown in Table I.

TABLE I. INFORMATION ABOUT HIS

Name	Resolution	# of pixels	# of classes	# of bands
Indian Pines	145x145	21025	16	200
Salinas	512x217	111104	16	204
Pavia University	610x340	207400	9	103

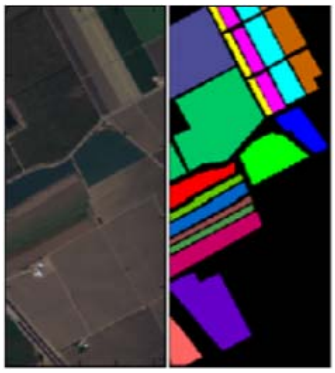
Indian Pines and Salinas datasets were collected by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor. Examples of images Indian Pines and Salinas with their respective GTs are shown in Figs. 3 and 4.

Originally the Indian Pines image had 220 bands and Salinas 224, but 20 water absorption bands were removed. The spatial resolution of these images is 20 m/pixel and the wavelength range from 0.4 to 2.5 μm. The classes of these images represent mainly crops characterizing themselves as agricultural areas. Both images have 16 classes represented by different colors in the GTs, and some Indian Pines classes represent Alfalfa, Corn, Grass, Oats, Soybean, Woods, among others. Salinas classes represent Broccoli, Corn, Fallow, Lettuce, Vineyard, among others.



(a) (b)

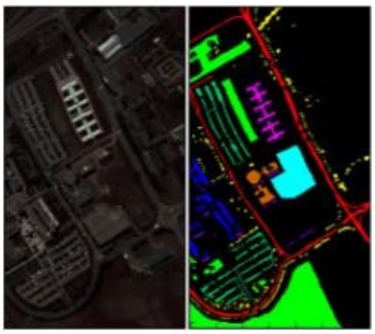
Figure 3. (a) Indian Pines HI. (b) GT.



(a) (b)

Figure 4. (a) Salinas HI. (b) GT.

The Reflective Optics System Imaging Spectrometer (ROSIS) obtained Pavia University dataset (shown in Fig. 5) with a spatial resolution of 1.3 m/pixel and their classes represent urban areas, and it has 9 classes, some of them being Water, Trees, Asphalt, among others.



(a) (b)

Figure 5. (a) Pavia University HI. (b) GT.

Six different band selection methods/cases were used for comparison:

i) GA-SVM [25];
ii) IRMoBS [29];
iii) WaLuMi [20];
iv-) MIMR-CSA [21];
v) IG-GWO [26];
vi) and HIs with all bands.

To avoid the influence of the mean filter on the pixels used in IDMMoBS, always the lower right pixel was used as a sample (systematic sampling). For each band, the mean value was obtained from the previous three pixels using the 2x2 mask. An example of the sampled pixels is shown in Fig. 6, where the black background pixel was used as a sample for IDMMoBS and the four pixels with the same number were used to calculate their mean value in each band.

1	1	2	2	3	3	4	4
1	**1**	2	**2**	3	**3**	4	**4**
5	5	6	6	7	7	8	8
5	**5**	6	**6**	7	**7**	8	**8**
9	9	10	10	11	11	12	12
9	**9**	10	**10**	11	**11**	12	**12**

Figure 6. Example of pixels used as samples (with black background) in experiments with the IDMMoBS.

For IDMMoBS, 20% of the sample pixels obtained by the previous method, that is, using black background pixel, were used for testing with 5-Fold Cross-Validation. For all other methods compared, 20% of all pixels of each HI were used for testing with 5-Fold Cross-Validation. An important observation is that the special information was considered only for IDMMoBS as part of the purpose of this paper, and this occurred both in the band selection process as well as in testing. For the other methods only the spectral information proposed in the original works was used.

The experiments with the comparative methods were similar to those performed in [26], so the comparison of these methods was possible. Only for the GA-SVM and IRMoBS futher executions were necessary to adjust the HIs and amount of samples used. For the GA-SVM, the fitness function was set with 0.8 and 0.2 respectively for accuracy and the inverse of the number of bands. Other parameters of these two algorithms were the same as the original proposals in [25] and [29].

The IDMMoBS, IRMoBS and GA-SVM have been set to run for 500 generations with 100 solutions (subproblems) in the population. The SVM was applied for the calculation of the overall accuracy with 3-Fold Cross-Validation. These values were empirically selected based on the runtime of each generation on the tested hardware. The crossover rate and mutation were 50% and 5% respectively. Other parameters defined in the IDMMoBS were neighborhood size $NS = 20$, p, the threshold of the amount of bands, with 30 (Indian Pines and Salinas) or 20 (Pavia University) and the weights w_1 and w_2 were randomly defined for each i-th subproblem. In the experiments, 80% of each HI pixels different from 20% testing pixels were used in the process of band selection in IDMMoBS.

After the band selection performed by each method, a classification was performed on the 20% testing pixels by the SVM using 5-Fold Cross-Validation to evaluate the final result. The optimal parameter C of the RBF kernel was determined via 5-Fold Cross-Validation.

In the GA-SVM, IRMoBS and IDMMoBS methods the bands were obtained automatically and their quantities are shown in the experiments. In the IG-GWO method the original number of bands obtained in [26], that is, 30, 28 and 18 bands were respectively used for the Indian Pines, Salinas and Pavia U. images. In the other three selection methods, the number of bands selected is 30 for the Indian Pines and Salinas datasets and 20 for the Pavia University dataset.

Three metrics were used to evaluate the success rate of classification of pixels:

i) Overall accuracy (OA) refers to the number of correctly classified instances divided by the total number of testing samples;

ii) Average accuracy (AA) is a measure of the mean value of the classification accuracies of all classes;

iii) The kappa coefficient (KC) is a statistical measurement of consistency between the ground truth map and the final classification map [26].

In addition to these metrics the number of bands used in each method is also discussed.

B. Results

The results obtained by the different methods and for each considered HI are shown in Tables II-IV. The first column shows the name of each method and also in parentheses is shown the total amount of bands used. The three results from the following columns respectively represent the overall accuracy (OA), the average accuracy (AA) and the kappa coefficient (KC). Best results in these tables are highlighted in bold.

Table II refers to results obtained for Indian Pines image. This Table shows that bands selected by IDMMoBS outperform the other six cases for classification using the SVM classifier. The good results provided by the IDMMoBS are repeated for the different metrics evaluated, that is, OA, AA and KC. In particular, it is observed that even the GA-SVM method, which is biased towards searching for solutions that directly improve the pixel classification, has no classification results similar to those provided by the bands selected by IDMMoBS. A plausible explanation for the good results achieved by the IDMMoBS in this image is because of the use of spatial information obtained by the mean filter. The results are also superior to other state-of-the-art literature methods such as IG-GWO and also for the images that use of all spectral bands.

TABLE II. OA, AA AND KC OBTAINED FROM DATA WITH BANDS SELECTED FROM INDIAN PINES BY DIFFERENT METHODS AND CLASSIFIED BY THE SVM USING 5-FOLD CROSS-VALIDATION

Indian Pines			
Method	**OA(%)**	**AA(%)**	**KC(%)**
GA-SVM (131 bands)	86.8 ±1.5	81.0 ±2.1	84.9 ±1.7
IRMoBS (44 bands)	80.0 ±1.7	74.6 ±4.6	77.2 ± 1.9
WaLuMi (30 bands)	80.6 ±0.8	68.2 ±2.0	77.7 ±1.0
MIMR-CSA (30 bands)	84.5 ±0.9	79.4 ±1.8	82.3 ±1.0
IG-GWO (30 bands)	85.2 ±0.8	82.6 ±6.6	83.1 ±0.9
All Bands (200 bands)	84.9 ±1.5	75.2±4.0	82.8 ±1.7
IDMMoBS **(29 bands)**	**91.4 ±1.1**	**91.4 ±3.9**	**90.1 ±1.3**

TABLE III. OA, AA AND KC OBTAINED FROM DATA WITH BANDS SELECTED FROM SALINAS BY DIFFERENT METHODS AND CLASSIFIED BY THE SVM USING 5-FOLD CROSS-VALIDATION

Salinas			
Method	**OA(%)**	**AA(%)**	**KC(%)**
GA-SVM (38 bands)	92.4 ±0.3	96.2 ±0.2	91.5 ±0.4
IRMoBS (32 bands)	88.3 ±0.6	93.5 ±0.5	86.9 ±0.6
WaLuMi (30 bands)	93.0 ±0.2	96.4 ±0.2	92.2 ±0.2
MIMR-CSA (30 bands)	93.5 ±0.2	96.8 ±0.1	92.7 ±0.2
IG-GWO **(28 bands)**	93.9 ±0.2	97.0 ±0.7	93.2 ±0.2
All Bands (204 bands)	92.8 ±0.3	96.5±0.2	92.1 ±0.4
IDMMoBS (29 bands)	**94.4 ±0.4**	**97.3±0.1**	**93.7 ±0.4**

TABLE IV. OA, AA AND KC OBTAINED FROM DATA WITH BANDS SELECTED FROM PAVIA UNIVERSITY BY DIFFERENT METHODS AND CLASSIFIED BY THE SVM USING 5-FOLD CROSS-VALIDATION

Pavia University			
Method	**OA(%)**	**AA(%)**	**KC(%)**
GA-SVM (70 bands)	93.9 ±0.5	92.9 ±1.1	92.0 ±0.7
IRMoBS (24 bands)	89.5 ±0.3	88.9 ±1.0	86.2 ±0.5
WaLuMi (20 bands)	90.1 ±1.0	86.7 ±1.2	86.7 ±1.4
MIMR-CSA (20 bands)	92.5 ±0.3	89.1 ±0.6	90.1 ±0.3
IG-GWO (18 bands)	94.2 ±0.2	92.3 ±1.0	92.3 ±0.2
All Bands (103 bands)	94.1±0.3	92.8±0.9	92.2±0.4
IDMMoBS **(15 bands)**	**96.2 ±0.3**	**95.5±0.5**	**94.7 ±0.4**

Tables III and IV refer to Salinas and the Pavia University images respectively. For these images most methods performed well for classification, often with metrics greater than 90%. These tables show that the IDMMoBS also presented better results when compared with the other methods in the classification criterion. The most competitive method to IDMMoBS was the IG-GWO which for Salinas and Pavia University images presented very close results.

In addition to comparing the OA, AA, and KC metrics, the mean value of the last solution found by each method and used to generate the data from Tables II-IV are shown in Table V. This Table shows the data on the number of bands reduction percentage for each evaluated method. These data can be calculated from the number of bands shown in the first column of Tables II, III and IV.

As shown in Table V, IDMMoBS also showed better results in the number of bands of the Indian Pines and Pavia University images, but for the Salinas image the IG-GWO performed better.

TABLE V. PERCENTAGE REDUCTION OF THE NUMBER OF BANDS DEFINED OR OBTAINED IN EACH METHOD.

Method	**Indian Pines**	**Salinas**	**Pavia University**
GA-SVM	34.5%	81.4%	32.0%
IRMoBS	78%	84.3%	76.7%
WaLuMi	85%	85.3%	80.6%
MIMR-CSA	85%	85.3%	80.6%
IG-GWO	85%	**86.3%**	82.5%
IDMMoBS	**85.5%**	85.8%	**85.4%**

By analyzing data classification and number of bands shown in the previous tables, in many methods simultaneously reducing the number of bands can lead to improved classification metrics. This behavior is present mainly when the IDMMoBS and IG-GWO methods are applied, which for many of the cases presented better results than the use of all bands and can suggest some improvement of the Hughes phenomenon.

IV. CONCLUSION

In this study, a novel supervised MOBS method called IDMMoBS was proposed. This method uses evolutionary algorithms and it is based on TDS, SVM, a repair strategy, spatial information and a centroid-based DM. The IDMMoBS searches for tradeoff solutions with the balance between the overall accuracy and the number of bands.

Experiments were conducted on HIs with different regions and diverse areas. HIs with selected bands by the proposed method were compared to pixel classification of original HIs with all bands and five other band selection methods. For most of the test cases conducted, the IDMMoBS, showed higher performance for the number of bands and better overall accuracy when applying the SVM classifier. In just one case, the method called IG-GWO achieved a slightly lower number of bands than the proposed method, but this method had a worse classification rate. Based on the tests performed the IDMMoBS has been able to find significant tradeoff solutions for specific classifiers and appear as a good alternative MOBS of HIs. IDMMoBS may have promoted good results in the classifier, mainly due to the use of spatial information incorporated in the band selection process and subsequent classification.

In future works, the band selection process will be evaluated with other algorithms such as decision trees and deep learning. Different parameters will also be tested, such as the number of generations, crossing rate, mutation, and

fitness functions. Other filter methods will also be considered to improve segmentation.

REFERENCES

[1] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, G. Trianni, "Recent advances in techniques for hyperspectral image processing", Remote Sensing of Environment, vol. 113, pp S110-S122, 2009. doi:10.1016/j.rse.2007.07.028.

[2] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, A. Abbas, "Modern Trends in Hyperspectral Image Analysis: A Review," IEEE Access, vol. 6, pp. 14118-14129, 2018. doi: 10.1109/ACCESS.2018.2812999.

[3] M. Attas, E. Cloutis, C. Collins, D. Goltz, C. Majzels, J. R. Mansfield, H. H. Mantsch,"Near-infrared spectroscopic imaging in art conservation: investigation of drawing constituents", Journal of Cultural Heritage, vol. 4, n. 2, pp. 127-136, 2003. doi:10.1016/S1296-2074(03)00024-4.

[4] M. Gong, M. Zhang and Y. Yuan, "Unsupervised Band Selection Based on Evolutionary Multiobjective Optimization for Hyperspectral Images," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 1, pp. 544-557, 2016. doi: 10.1109/TGRS.2015.2461653

[5] T. M. Lillesand, R. W. Kiefer, J. W. Chipman, "Remote Sensing and Image Interpretation", pp. 550-562, 5th ed., Ed. John Wiley & Sons, 2004. doi: 10.2307/634969.

[6] S. Amini, S. Homayouni, A. Safari, A. A. Darvishsefat, "Object-based classification of hyperspectral data using Random Forest algorithm", Geo-spatial Information Science, vol. 21, pp. 127-138, 2018. doi: 10.1080/10095020.2017.1399674.

[7] J. Xia, P. Ghamisi, N. Yokoya, A. Iwasaki, "Random Forest Ensembles and Extended Multi-Extinction Profiles for Hyperspectral Image Classification." IEEE Transactions on Geoscience and Remote Sensing, vol. 56, n. 1, pp. 202-2016, 2018. doi: 10.1109/TGRS.2017.2744662.

[8] D. A. Landgrebe, "Signal Theory Methods in Multispectral Remote Sensing", John Wiley and Sons, pp. 237-239, 2003. doi: 10.1002/0471723800.

[9] G. Hughes, "On the mean accuracy of statistical pattern recognizers". IEEE Trans. Inf. Theory, vol. 14, n. 1, pp. 55–63, Jan. 1968. doi: 10.1109/TIT.1968.1054102.

[10] X. Zhang, Q. Sun, J. Li, "Optimal band selection for high dimensional remote sensing data using genetic algorithm", Proceedings of SPIE - The International Society for Optical Engineering., October 2009. doi: 10.1117/12.847907.

[11] D. Saqui, J. H. Saito, L. A. C. Jorge, E. J. Ferreira, D. C. Lima, J. P. Herrera, "Methodology for band selection of Hyperspectral Images using Genetic Algorithms and Gaussian Maximum Likelihood Classifier", International Conference on Computational Science and Computational Intelligence, Las Vegas, EUA, pp. 733-738, 2016. doi: 10.1109/CSCI.2016.0143.

[12] M. Kumar, "Feature Selection for Classification of Hyperspectral Remotely Sensed data using NSGA-II", Water Resources Seminar, Citeseer, 2004.

[13] M. D. Farrell, R. M. Mersereau, "On the impact of PCA dimension reduction for hyperspectral detection of difficult targets", IEEE Geoscience and Remote Sensing Letters, vol. 2, no. 2, pp. 192–195, Apr. 2005. doi: 10.1109/LGRS.2005.846011.

[14] N. Falco, J. A. Benediktsson, L. Bruzzone, "A study on the effectiveness of different independent component analysis algorithms for hyperspectral image classification", IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 7, no. 6, pp. 2183–2199, Jun. 2014. doi: 10.1109/JSTARS.2014.2329792.

[15] M. Masaeli, G. Fung, J. G. Dy, "From transformation-based dimensionality reduction to feature selection". ICML'10 Proceedings of the 27th International Conference on International Conference on Machine Learning, pp. 751 – 758, 2010.

[16] K. Sun, X. Geng, L. Ji., "Exemplar component analysis: A fast band selection method for hyperspectral imagery", IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 5, pp. 998–1002, May 2015. doi: 10.1109/LGRS.2014.2372071.

[17] W. Sun, L. Zhang, B. Du, W. Li, Y. M. Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 6, pp. 2784–2797, Jun. 2015. doi: 10.1109/JSTARS.2015.2417156.

[18] W. Sun, L. Zhang, L. Zhang, Y. M. Lai. "A dissimilarity-weighted sparse self-representation method for band selection in hyperspectral imagery classification," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 9, no. 9, pp. 4374–4388, Sep. 2016. doi: 10.1109/JSTARS.2016.2539981.

[19] G. Zhu, Y. Huang, J. Lei, Z. Bi, F. Xu., "Unsupervised hyperspectral band selection by dominant set extraction," IEEE Geoscience and Remote Sensing Letters, vol. 54, no. 1, pp. 227–239, Jan. 2016. doi: 10.1109/TGRS.2015.2453362.

[20] A. Martínez-Usó, F. Pla, J. M. Sotoca, P. García-Sevilla, "Clustering-Based Hyperspectral Band Selection Using Information Measures". IEEE Transaction Geoscience Remote Sensing 2007, vol. 45, pp. 4158–4171. doi 10.1109/TGRS.2007.904951.

[21] J. Feng, L. Jiao, F. Liu, T. Sun, X. Zhang, "Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images". Pattern Recognition 2016, 51, 295–309. doi: /10.1016/j.patcog.2015.08.018.

[22] R. Y. M. Nakamura, L. M. G. Fonseca, J. A. dos Santos, R. da S. Torres, X.-S. Yang, J. P. Papa, "Nature-inspired framework for hyperspectral band selection," IEEE Trans. Geoscience Remote Sensing, vol. 52, no. 4, pp. 2126–2137, Apr. 2014. doi: 10.1109/TGRS.2013.2258351.

[23] H. Su, B. Yong, Q. Du, "Hyperspectral band selection using improved firefly algorithm," IEEE Geosci. Remote Sens. Lett., vol. 13, no. 1, pp. 68–72, Jan. 2016. doi: 10.1109/LGRS.2015.2497085.

[24] C. Vaiphasa, A. K. Skidmore, W. F. Boer, T. Vaiphasa. "A hyperspectral band selector for plant species discrimination". ISPRS Journal of Photogrammetry and Remote Sensing, vol. 62, n. 3, p. 225-235, 2007. doi: 10.1016/j.isprsjprs.2007.05.006

[25] L. Zhuo, J. Zheng, F. Wang, X. Li, B. Ai, J. Qian, "A Genetic Algorithm Based Wrapper Feature Selection Method for Classification of Hyperspectral Images Using Support Vector Machine". Proceedings of SPIE 7147, Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Classification of Remote Sensing Images, vol. 7147, pp. 397–402, 2008. doi: 10.1117/12.813256.

[26] X. Zhang, W. Wang, Y. Li, L. C. Jiao, "Pso-based automatic relevance determination and feature selection system for hyperspectral image classification", Electronics Letters, vol. 48, n. 20, pp. 1263–1265, 2012. doi: 10.1049/el.2012.0539.

[27] M. Zhang, J. Ma, M. Gong. "Unsupervised Hyperspectral Band Selection by Fuzzy Clustering with Particle Swarm Optimization". IEEE Geoscience and Remote Sensing Letters, vol. 14, nº5, pp. 773-777. 2017. doi: 10.1109/LGRS.2017.2681118.

[28] F. Xie, F. Li, C. Lei, L. Ke, "Representative Band Selection for Hyperspectral Image Classification". ISPRS International Journal of Geo-Information 2018, 7, 338. doi:10.3390/ijgi7090338.

[29] X. Xu, Z. Shi, B. Pan, "A New Unsupervised Hyperspectral Band Selection Method Based on Multiobjective Optimization," in IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 11, pp. 2112-2116, Nov. 2017. doi: 10.1109/LGRS.2017.2753237.

[30] N. Sánchez-Maroño, A. Alonso-Betanzos, M. Tombilla-Sanromán, "Filter methods for feature selection - A comparative study". International Conference on Intelligent Data Engineering and Automated Learning. Lecture Notes in Computer Science, vol. 4881, Springer, Berlin, Heidelberg, 2007. doi:10.1007/978-3-540-77226-29.

[31] X. Cao, B. Ji, Y. Ji, L. Wang, L. Jiao, "Hyperspectral image classification based on filtering: A comparative study". Journal of Applied Remote Sensing, vol. 11. doi: 10.1117/1.JRS.11.035007.

[32] Z. Qingfu, Li, Hui, "MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition". IEEE Transactions on Evolutionary Computation, Vol. 11, N. 6, 2007. doi: 10.1109/TEVC.2007.892759.

[33] D. Kimovski, R. Matha, S. Ristoy, R. Prodan, "Multiobjective service oriented network provisioning in ultra-scale systems". European Conference on Parallel Processing, Lecture Notes in Computer Science, vol 10659. Springer, Cham pp. 529-540, 2017. doi: 10.1007/978-3-319-75178-8_43.

[34] V. N. Vapnik, "The Nature of Statistical Learning Theory". New York: Springer-Verlag, 1995. doi: 10.1007/978-1-4757-3264-1.

[35] B. Boser, I. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in Proc. 5th Annu. Workshop Comput. Learn. Theory, 1992, pp. 144–152. doi: 10.1145/130385.130401.

[36] N. Cristianini and J. Shawe-Taylor. Cambridge, U.K.: Cambridge Univ. Press, 2000. doi: 10.1017/CBO9780511801389.

[37] M. Pal and G. M. Foody, "Feature Selection for Classification of Hyperspectral Data by SVM," in IEEE Transactions on Geoscience and Remote Sensing, vol. 48, no. 5, pp. 2297-2307, May 2010. doi: 10.1109/TGRS.2009.2039484.

[38] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," IEEE Geosci.Remote Sens. Mag., vol. 5, no. 1, pp. 8–32, Mar. 2017. doi: 10.1109/MGRS.2016.2616418.