



AKADÉMIAI KIADÓ



International Review of
Applied Sciences and
Engineering

12 (2021) 1, 64–75

DOI:

[10.1556/1848.2020.00149](https://doi.org/10.1556/1848.2020.00149)

© 2020 The Author(s)

ORIGINAL RESEARCH
PAPER



Reduction dimensionality of hyperspectral imagery using genetic algorithm and mutual information and normalized mutual information as a fitness function

Merzouqi Maria*, Sarhrouni El Kebir and Hammouch Ahmed

Electronic Systems, Sensors and Nanobiotechnologies (E2SN), ENSET Mohammed V, University Rabat, Salé, Morocco

Received: July 03, 2020 • Accepted: August 06, 2020

Published online: February 10, 2021

ABSTRACT

Hyperspectral images (HSI) present a wealth of information. It is distinguished by its high dimensionality. It served humanity in many fields. The quantity of HSI information represents a double-edged sword. As a consequence, their dimensionality must be reduced. Nowadays, several methods are proposed to overcome their duress. The most useful and essential solution is selection approaches of hyperspectral bands to analyze it quickly. Our work suggests a novel method to achieve this selection: we introduce a Genetic Algorithm (GA) based on mutual information (MI) and Normalized Mutual Information (NMI) as fitness functions. It selects the relevant bands from noisiest and redundant ones that don't contain any additional information. The proposed method is applied to three different HSI: INDIAN PINE, PAVIA, and SALINAS. The introduced algorithm provides a remarkable efficiency on the accuracy of the classification, in front of other statistical methods: the Bhattacharyya coefficient as well as the inter-bands correlation (Pearson correlation). We conclude that the measure of information (MI, NMI) provides more efficiency as a fitness function for GA selection applied to HSI; it must be more investigated.

KEYWORDS

reduction of dimensionality, mutual information, normalized mutual information, genetic algorithm, SVM

1. INTRODUCTION

Due to the excellent information provided by the hyperspectral image, it becomes a needed tool in several areas: such as target detection, land covers surveys, and mineralogy. Besides their advantages like the good precision provided by the dense sampling of the spectral signatures of the terrestrial coverings; it also has drawbacks that lie in their weight. They are usually made up of hundreds of close spectral bands captured from the same scene. These bands are combined to produce three-dimensional data called a hyperspectral data cube [1].

1.1. The necessity of dimensionality reduction of HSI

The bulky information of Hyperspectral images (HSI) has a negative effect on the calculation time and the accuracy of the classification also Hughes demonstrated that as the number of samples increases the classification rate decreases, this phenomenon named in literature by the curse of dimensionality Hughes [2]. HSI contains beside the informative bands other irrelevant and redundant. It is clear to delete these bands which haven't any useful information. Despite that, it has a negative impact on the performance of these images. Several methods have been proposed in previous years to treat this problem without losing useful

*Corresponding author.

E-mail: maria.merzouqi@um5s.net.ma

information. Among them the dimension reduction technique, which is divided into two main categories: feature selection and feature extraction.

The extraction Approach of characteristics [3, 4] is used to reduce the large space by a linear or nonlinear transformation to low-dimensional space, but it cannot retain the original physical interpretation of the image. Whereas the spectral distance between the neighboring bands is only 10 nm. Besides that, the correlation between them is so high. All this tends towards the non-recommendation of feature extraction [5, 6] for the dimensionality reduction of HSI.

The selection features are much recommended than the extraction of characteristics [7–9]. It's able to overcome the limitations of the previous method, with high calculation accuracy and classification efficiency without distorting the original data, since no transformation is introduced. The selection of attributes is done in two main steps. It is used to calculate, often scores, or even distances while respecting a defined strategy. There are three categories of strategies approved in the literature: wrappers, filters, embedded [10–12].

The filtering strategy [13] was also used as a pre-processing step for the strategy wrapper [14]. This often used. It is renowned for its speed and precision. It acts quickly since it's based on the optimization of a defined function, while the wrapper is heavier. It uses a predictive model respecting the minimization of an error rate to evaluate the functionality of each selected attribute. Each decision of selection calls the classifier that makes the process of reduction heavy. The wrapper is not feasible for real-time processing. The filter is more useful for processing sensitive to latency. Although there is another strategy that combines the advantages of the two methods already mentioned: briefly, the learning part and the function selection part are inseparable in this strategy named embedded [11, 12].

However, each approach has its potential. Several optimization methods have been examined in the context of the reduction of hyperspectral dimensionality. Filtering methods like Fisher's filter [15, 16], calculated by [17] they propose a generalized Fisher score to jointly select features, to find the subset of features that maximizes the lower bound of traditional Fisher score.

In our proposition, we adopt a filter strategy as described in the next section.

1.2. Mutual information as a criterion for band selection

Mutual information approach (MI) [18] has several uses it can be a selection criterion between two variables or more of the membership. While the evaluation is done independently and the final selection is done according to the principle: the entity that maximizes MI will be taken [19–22].

The MI has been used by [23–25] for the selection of HSI based on the principle of fusion images, which relies on the creation of a new ground truth estimated. As the first step:

the MI calculated between each band and the real reference. The chosen bands to build an estimated reference must have a high value of MI. As a second principle, they compare the two MI for the real and estimated reference. Besides, the method doesn't require prior knowledge, which made it more powerful.

(Yingxuan Zhu, Pramod K. Varshney, Hao Chen 2007), propose a method based on independent component analysis ICA [26], as well as the correlation coefficient and the MI. They merge the information of numerous bands into three fit images for color billposting. For the evaluation of the performances of their method, they introduced the entropy and the edge correlation.

In the proposed method we use MI, normalized mutual information (NMI) as to fitness function for the genetic algorithm (GA) and we compare it with, Correlation coefficient, and Bhattacharyya coefficient.

1.3. Previous uses of GA for dimensionality reduction

Reduction methods are often based on several metrics. They have shown their potential on HSI. Taking the case of [27] used the Bhattacharyya distance as a criterion function with GA-IDA to reduce the hyperspectral image adopting the selection process. The GA approach was introduced as a filter strategy [26, 28]. Its fitness function maximizes the Bhattacharyya distance between the distributions of two probabilities. Also, the IDA application had two main entries (the fisherman's metric report and the selected subset). It has a positive contribution to the classification rate.

As we already know, reducing the size of HSI is an essential step to ensure proper use. This reduction is made according to different methods, namely the identification, the selection, the extraction of the relevant elements from a large set of unsorted elements. Each method has its complexity in the application [29]. On the other hand, they have the greatness of efficiency in the selection.

According to the aforementioned studies, GAs, as the name suggests, come from a source in the natural sciences, they are made up of biological mechanisms, namely selection, mutation, and crossing. John Holland introduced GAs in the early 1970s [30]; they have managed to penetrate all areas, thanks to their efficiency in solving all kinds of problems among them the curse of dimensionality of HSI. They already have a relatively long history, since John Holland's first work on adaptive systems dates back to 1962 [31]. The GA is very popular as an effective tool in solving large-scale data limitations because it interacts following a process of natural evolution. GA has been successfully applied in neural networks, machine learning, and text categorization technology [32, 33]. It operates according to the objectives sought, whether the reduction in dimensionality or the reduction of noise, the elimination of redundant information, increased precision for a given forecasting task (selection, extraction).

We will not go further than the previous studies. We will adopt the same operating principle. In this work, the application of the GA will be different. Since the objective



and the dataset on which this method is applied are different. Also, to evaluate the proposed method we used two comparison fitness functions which are presented in the following sections.

GA is a flexible algorithm that adapts easily to changing situations and data. This makes this evolutionary approach adequate for our case.

The GA-SVM Models [34, 35] have been used for a long time on several domains for different purposes.

Its presence in our days confirms their effectiveness in scientific research, namely remote-sensing since it achieves the desired objective, such as the selection and classification of HSI, with great precision in short calculation time compared to other sequential algorithms.

Our proposal is based in particular on GA-SVM that depends on the MI and NMI, as the fitness function. It is a selection filter to optimize the size of the hyperspectral image; the established fitness function is the measure of the MI dependence between bands to eliminate irrelevant and redundant bands. We compare the results with the following metrics: the Bhattacharyya coefficient, the static correlation coefficient.

The rest of this article will first be devoted to detailing the proposed method of dimensionality reduction with the principle of band selection by GA-MI-SVM/GA-NMI-SVM. Section 3 describes the experimental data set used in this work. Section 4 presents the results of the classification of the algorithms. On the other hand, the analysis and the conclusion will be established in the last part while quoting the perspectives of the proposed study.

2. METHODOLOGY USED

2.1. Genetic algorithm

Nowadays, GAs are the most common among evolutionary algorithms to overcome this type of obstacle. This allows us to find effective potential solutions with a quick response time. GA is a heuristic technique based on the same biological principle but applied numerically, such that individuals represent a chromosome embodying a set of bits [0 1]. A population of chromosomes is a set of individuals. The same biological steps will be applied to the population without exception. The chromosome survived during each generation will be evaluated by a fitness function that aims to select the best chromosome meeting the need.

In our case, the initial population will be chosen randomly. The evolutionary change that would be established on the population is well determined. At this time, the different genetic operations, namely: crossing, mutation, and selection will be applied to the individuals of this population for the development and reproduction of another population, which must meet the requirement of the fitness function [36–39].

The appropriate model of our case of band selection will be detailed as follows: each chromosome contains some genes B equal to the size of the bands, and each gene

represents a random number either 0 or 1, associated with a band. As already mentioned, the quality of each chromosome is evaluated according to a fitness function that serves to measure the dependency between the bands. The classification of the global accuracy (OA), as well as the average accuracy (AA), KAPPA), performed by an SVM-RBF, will be applied on the chromosome selected. In addition, GA operators were used is: crossover = 0.98, as well as mutations = 0.005.

Selection tournament = the selection of tournament size is the main purpose of organizing several tournaments between individuals, from which the choice of some individuals (or chromosomes) is done randomly in the population. The qualified chromosomes will be chosen to move to the next step.

Elitism = this step aims to preserve the perfect chromosome while creating a copy of this chromosome in the new offspring population that could be lost in the next steps.

Crossover = or recombination is an important operator for the construction of offspring. It combines two partial parts of elements from two parents selected randomly in the previous step to create two new children. They are considered as one of the solutions proposed by the population involved (Fig. 1).

Mutation = is another operator that ensures the diversity of populations from one generation to another. In this step, one or more of the values of the genes in the chromosomes selected will be modified relating to its origin. This operation is established according to a prior definition of a low probability by the user (Fig. 2).

Fitness function: it is a function that has a decisive role regarding the report (solution problem). In other words, it defines whether the candidates overcome the limitations presented. It's a fast and repetitive calculation function for each individual in each generation.

2.2. The measure of dependence: mutual information

The purpose of this study is to select the features that contain more information regarding the output and remove the one that is irrelevant and redundant. For this reason, we have introduced the most robust measure of MI as a

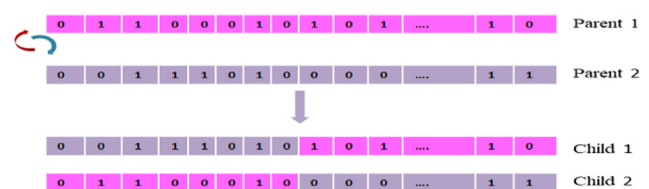


Fig.1. Crossover operator

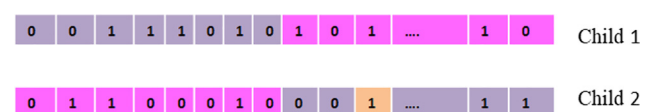


Fig. 2. Mutation operator



selection criterion [40]. MI is statistical measure it based on the concept of Shannon's information theory [41] to determine the relationship between variables.

$$I(A, B) = \sum_{a \in A} \sum_{b \in B} P(a, b) \log_2 \frac{P(a, b)}{P(a)P(b)}. \quad (1)$$

The MI and the entropy have several relationships, for example:

$$I(A, B) = H(A) + H(B) - H(A, B). \quad (2)$$

(A, B) respectively represent the selected bands and the ground truth. A great value of the MI between a pair of variables confirms the dependence between them. $H(A)$ is the entropy of Shannon [41]. It measures the uncertainty of a discrete random variable A. with the $P(a)$, the probability of A is defined as follows:

$$H(A) = -\sum_{a \in A} P(a) \log P(a). \quad (3)$$

In the literature, the NMI has several forms [42, 43, 56] (Fig. 3).

(Studholme, 1999) used NMI, Eq. (4), to measure similarity. This measure proved its reliability on the 3D medical images. It is defined as follows:

$$NMI = \frac{H(A) + H(B)}{H(A, B)}. \quad (4)$$

$H(A, B)$ is the joint entropy between the selected bands and the ground truth. The candidate band must maximize this measure to be selected [7].

3. THE HYBRID ALGORITHM PROPOSED

The first generation of the binary population is chosen randomly. The length of each chromosome equals the number of bands for each image it contains a series of [0 = unspecified band, 1 = chosen band]. In this way, for each chromosome, the options of the GAs will be applied, and then the fitness function that represents the MI [44, 45] or NMI [37, 46, 47] will be calculated for each gene in each chosen chromosome for several generations. One or more

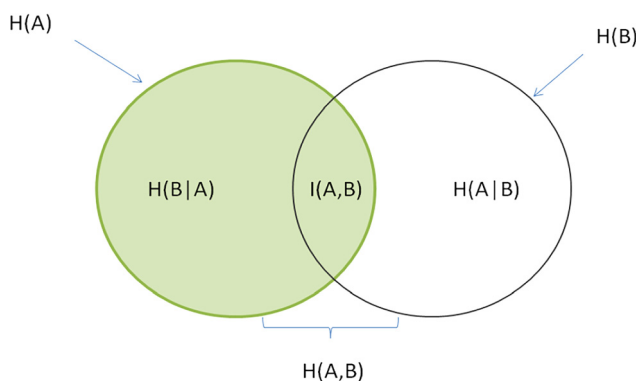


Fig.3. Illustration of the relationship between mutual information and entropy

MI [13] values of the selected chromosome will be obtained. The chromosome that maximizes the fitness functions quoted in the next section will be selected as the final result. The following flowchart illustrates the proposed hybrid approach.

The metrics Overall Accuracy (OA), Average Accuracy (AA), Kappa Coefficient (K) are widely used in hyperspectral image processing. Here in this study, we will calculate them. The result of the proposed method will be compared to the other results of two different fitness functions: Correlation coefficient [48–50], and Bhattacharyya coefficient [51, 52]. The proposed fitness function for the GA will be used for the first time and applied to the hyperspectral image processing. Our goal of this diversity is to demonstrate the strength of our method (Fig. 4).

3.1. Fitness functions for comparison

✓ **Bhattacharyya coefficient or distance:** measure the similarity between two probability distributions. A small value of this measure means a true similarity. The

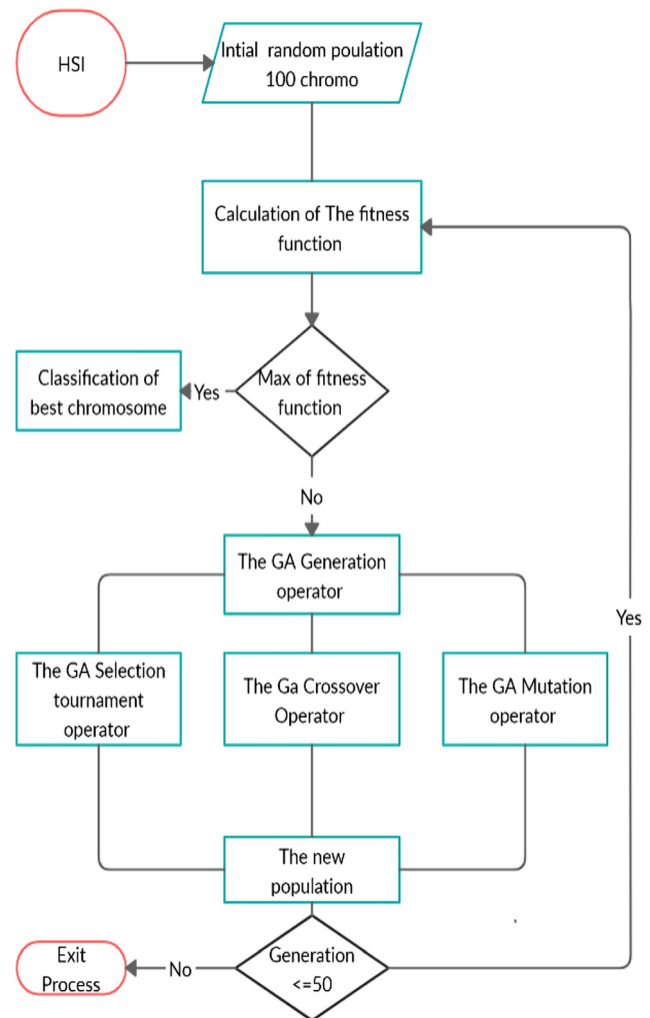


Fig. 4. The Flowchart of the proposed method

Bhattacharyya distance is only the natural logarithm of the Bhattacharyya coefficient, which is used to measure the overlap between two distributions. They are relatively related, the increase of the Bhattacharyya coefficient implies the decrease in the distance. In this work, the chosen band must maximize the coefficient. In our study, we used the Bhattacharyya coefficient as a fitness function for a logical comparison with the other scores proposed for this method.

$$BC(P, P') = \sum_{a \in A, b \in B} \sqrt{P(a)P'(b)}. \quad (5)$$

$P(a)$ and $P'(b)$ is the probability for each variable A and B.

✓ **Coefficient of correlation:** The correlation determines the statistical association between two variables. By the measure of dependence of two variables, it makes it possible to predict the exploitable relation in reality between them. The coefficient of correlation is the ratio between the product of the covariance of the two variables and the product of their standard deviations (respectively).

$$\text{Corr}(A, B) = \frac{\text{cov}(A, B)}{\sigma_A \sigma_B} = \frac{E[(A - \mu_A)(B - \mu_B)]}{\sigma_A \sigma_B}. \quad (6)$$

$\text{cov}(A, B)$ is the covariance of A and B, σ_A σ_B standard deviations of A and B. So the correlation is only the normalization of the covariance which takes positive or negative values. This normalization is established by their multiplication with the standard deviation of the variables.

The correlation coefficient is independent of the units of measure of the variables, it is bounded by

$$-1 < \text{Corr} < +1.$$

- When $0 < \text{Corr} \leq 1$ the link between the two variables tends to be linear positively. That is to say, they are similar in the same way. Knowledge of X provides the value of Y (and vice versa).
- When $-1 < \text{Corr} < 0$, the bond is negatively linear. More precisely, the variables are highly correlated but in the opposite direction. Lastly, where $\text{Corr} = 0$ this means that the variables are strictly independent.

3.2. Classifier used: support vector machines-RBF/KNN

For the validation of our results, we carefully chose RBF-SVM (Support Vector Machines), one of the most used supervised classifiers in the field of remote sensing. This choice was made after testing the other kernel. The RBF kernel was able to achieve important results with an amenable number of learning samples [53]. SVM is a learning machine that analyzes the data used for classification. Its operating principle lies in the construction of an optimal hyperplane of two classes with a precise margin of separation, the distance between the class, and this margin must be large. It is also characterized by several kernels. In

our study, the parameters C and γ were defined by the five-fold cross-validation operation. This work is also evaluated by the KNN classifier [54]. This step aims to judge the choice of the SVM as classifier for the classification stage. In all experiments, 50% of pixels of classes are chosen randomly and labeled to constitute the training samples and the rest of the pixels are used for the test and validation. The best known metrics [55] in the classification domain stating the quantitative measure of the performance: Kappa coefficient (KA) and the classification rate: Overall accuracy (OA), average accuracy (AA) was computed.

- Kappa coefficient (KA) aims to measure the quality of the final classification map. Computing the consistency between the ground truth and the obtained map.
- Overall Accuracy (OA) is a percentage calculated by dividing the total number of the correct classification of an instance by the total number of test samples
- Average accuracy (AA) indicates the average value of the exact classification for all classes.

4. STUDY AREA

The compilation of our algorithm is done on the scientific programming language Matlab 2015 which is positively famous and often used for the image processing, on a computer with a core (TM) i7-6500U CPU and 2.50 GHz 2.60 GHz frequency, its operating system has 64 Kbits with 8 GB of RAM.

To validate the robustness of our proposed method we had applied our algorithms on 3 data sets with different contents.

We have used the AVIRIS sensor (Fig. 5), the Indiana Pine image was taken at a test site in northwestern Indiana

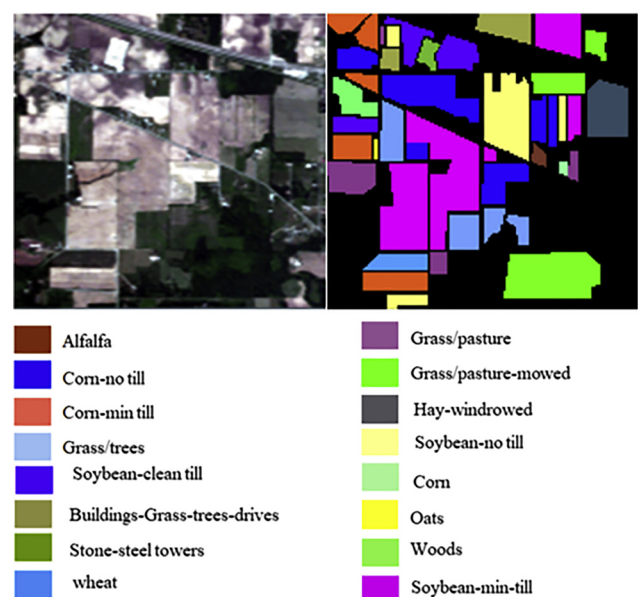


Fig. 5. The color composite and the corresponding ground truth with class labels for Indian

and has 220 spectral reflectance bands in a well-defined wavelength range: 0.4 and 2.5 μm . Each band consists of 145×145 pixels. This scene contains a diversity of agro-ecological and urban zones. It has two-thirds of agriculture and one-third of forests or other natural perennials. And for the urban area is the presence of motorways, railways, low-density housing, and other built structures and smaller roads. The basic truth provided, content 10,366 of pixels labeled with a number that represents the number of classes. It contains 16 classes. The Black zone represented by zeros and indicates pixels that are not labeled [16].

ROSIS Pavia is a hyperspectral image taken by the sensor ROSIS-03 (Reflective Optics Imaging Spectrometer) from the urban area of the engineering school of the University of Pavia in Italy. It contains 103 bands of 610×340 pixels that reflect the reflectance spectral wavelengths in the range of 0.43–0.86 μm . The sum of the pixels labeled by 1–9 classes in the ground truth provided is 42,776 [16] (Fig. 6).

This scene was collected by the AVIRIS sensor of the Salinas Valley, California, it has 224 bands. It is characterized by a high spatial resolution (3.7 m/pixel). The covered area includes 512 lines out of 217 samples, which determines vegetables, bare soil, and vineyard fields. The ground truth of Salinas contains 16 classes [16] (Fig. 7).

5. EXPERIMENTAL RESULTS AND DISCUSSION

5.1. Fitness functions for the three datasets.

The following curves (Fig. 8). represent the fitness functions used in this work of each image: Salinas and Indian Pine which contain 16 classes labeled, unlike Pavia which has just 9 classes.

✓ From the curves above, we deduce the bands that have low values of the MI and the NMI, and a few presences of the correlation, carries no information. See the first band of the images Pavia.

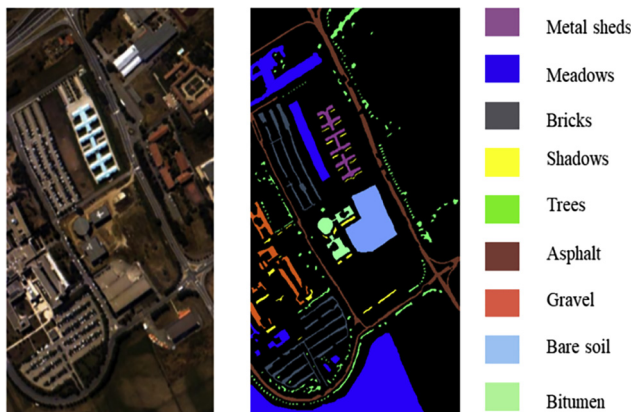


Fig. 6. The color composite and the corresponding ground truth with class labels for Pavia

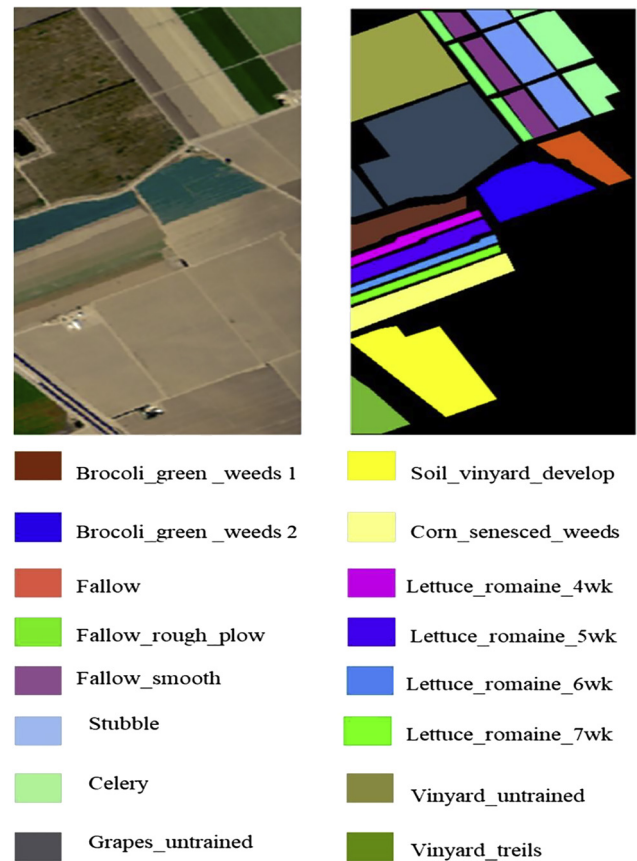


Fig. 7. The Color composite and the corresponding ground truth with class labels for Salinas

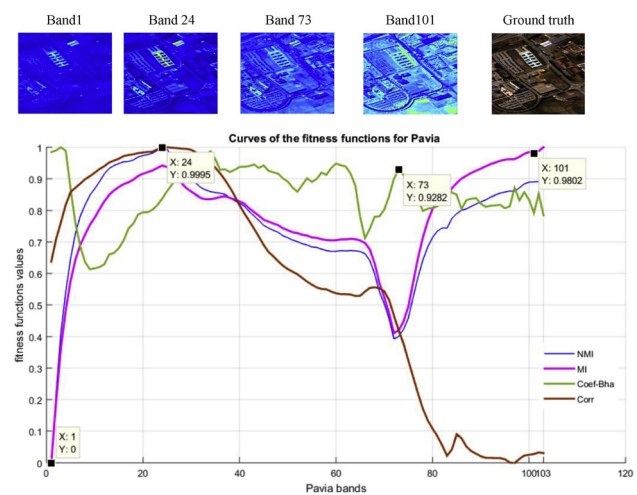


Fig. 8. The bands and curve illustration of all fitness function used of HSI Pavia

✓ The diversity of images influences the results of the proposed methods.

A band that has a meager value: either MI or normalized and uncorrelated, remains disfigured even with the strong existence of Coef-Bha. For example, band 155 for Salinas in Fig. 9.

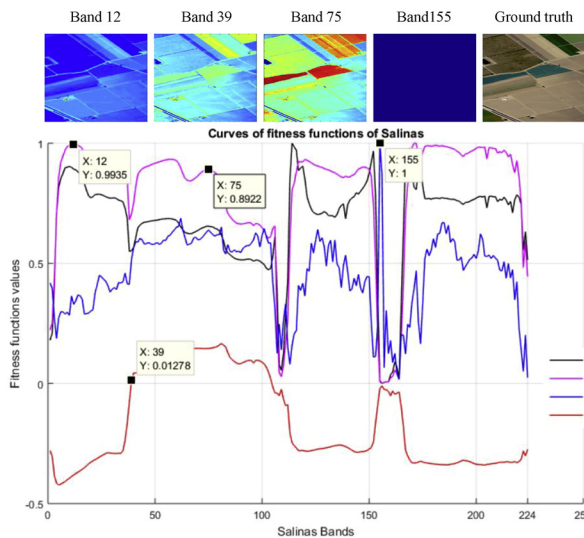


Fig. 9. The bands and curve illustration of all fitness functions (MI: Mutual Information, NMI: Normalized MI, Corr: Pearson correlation, Coeff-Bha: Bhattacharyya Coefficient) used of HSI Salinas

- ✓ The absence of each one: MI, NMI, the coefficient of Bhattacharyya also the correlation, indicates that the image has no added value. See band 151 in the following figure Fig. 10.
- ✓ From these curves, we deduce that the presence of fitness functions MI, NMI, Bhattacharyya coefficient and the correlation coefficient give high values for bands similar to the ground truth so they must be maximized in the algorithm

5.2. Result of the proposed algorithm

The following section is devoted to the presentation of the experimental results of our proposed method (AG-MI/NMI-SVM). It confirmed its robustness in front of the other fitness functions of the AG algorithms reproduced in our way. Furthermore, to prove that its discriminative selection is firm, another KNN classifier is introduced for the classification of the three images which contain a diversity of classes that had reduced values in front of the chosen SVM. The following Tables (1–3) represent the detailed results of a single set = 80 bands for Indiana and the Salinas which are composed of the same number of classes. As well, Table 3 shows the results of 36 Pavia bands comprising 9 classes. The columns indicate the values of the three metrics of the classification as well as the precision on of each class.

As an analysis of the following tables, it became clear that the proposed method marked its strong presence with its considerable values of the different metrics according to the two classifiers. Despite the weakening of the classification rate by KNN, which leaned its sensitivity against certain classes, namely the classes (Bare Soil for Pavia Hay-wind-rowed, Corn, Oats for Indiana Pine and Vinyard_untrained for Salinas) which are an agricultural aspect, and contrary to the other algorithms reproduced. Our method has been able to keep its effectiveness continuously in the discrimination of classes for the three images.

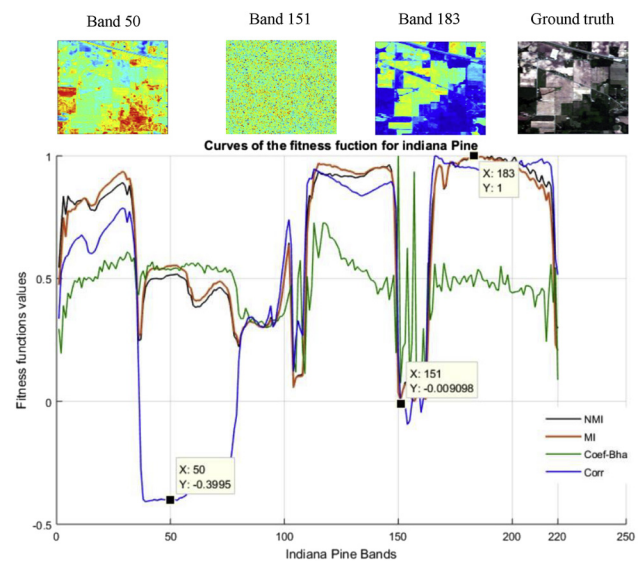


Fig. 10. The bands and curve illustration of all fitness functions (MI: mutual information, NMI: normalized MI, Corr: Pearson correlation, Coeff-Bha: Bhattacharyya coefficient) used of HSI Indiana Pine

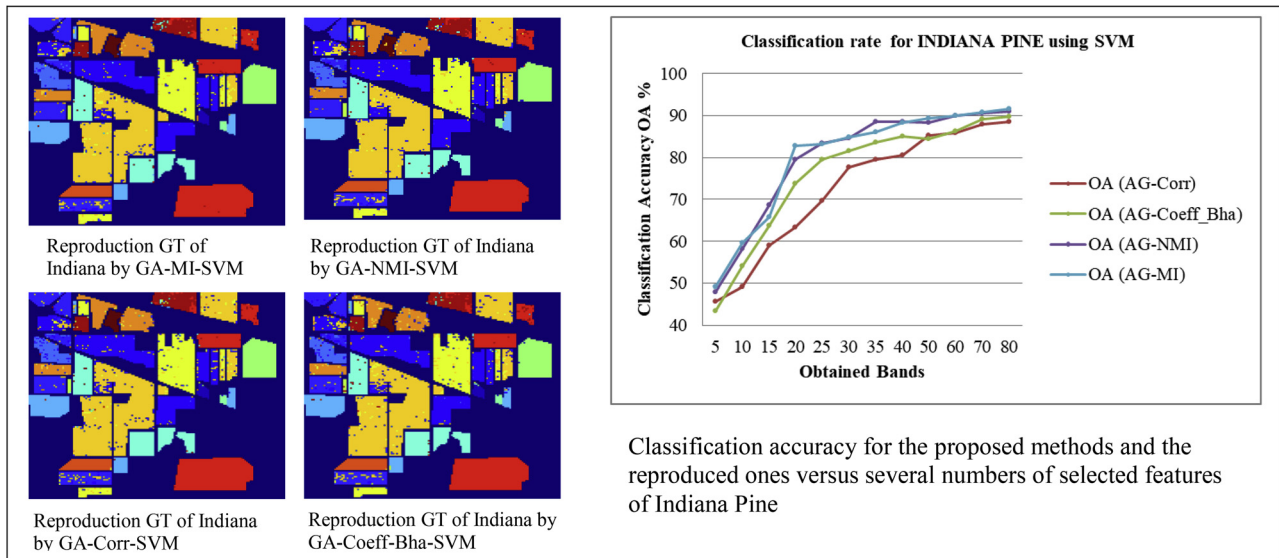
The inclusion of the curves and reproduced ground truth for the retained bands below aims to examine the behavior of the dimensionality reduction, as well as the compatibility of the classifier adopted in our method face to the other reproduced algorithms.

5.3. Analyses of results

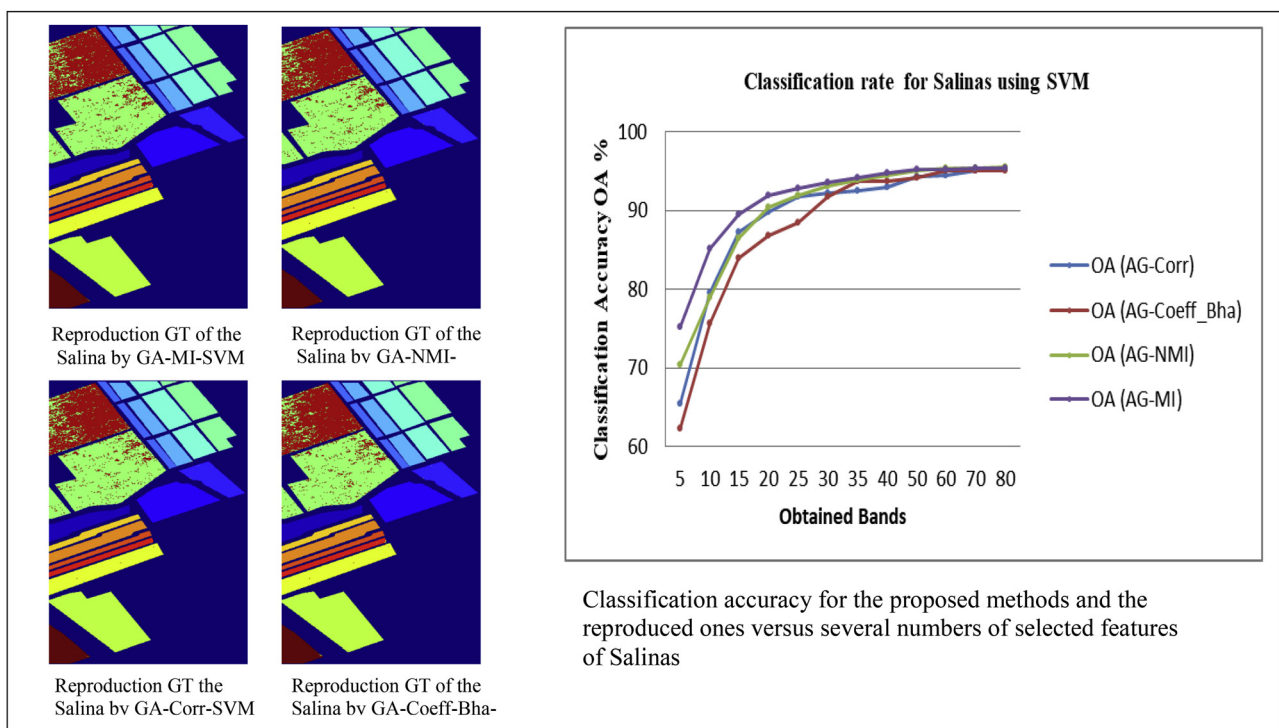
–For HIS Pavia: the fitness function based on the measurement of information (MI and NMI) gives high accuracy of classification OA with fewer bands than the other fitness functions (AG-coeff_bha, AG-Corr). The AG-NMI had OA = 90.04%, for only 15 bands of Pavia datasets. It can also reach more than 94% of the classification rate by SVM for 35 bands with precision OA = 94.43%, AA = 92.46%, KA = 92, 59%. The selection was hard and effective since it could not exceed 36 bands. We can deduce that NMI is an adequate fitness function to reduce the dimensionality of Pavia.

–MI is classified in the second-order using their classification rate compared to the other functions. AG-Corr is close to MI that gave important values with OA = 84.33% for 35 bands, and it has exceeded 94% for 65 bands when the AG-Corr obtained only OA = 93.42 when AG-Coeff-Bha classified the last one according to their values obtained OA = 93.07% for the same number of bands.

–The fitness function based on measurement information (MI and NMI) has once again proven its effectiveness in combating the high dimensionality of these images, as well as SALINAS: they reach precious accuracy values of the classification even for small sets of bands. They are therefore more favorable than other fitness. They made a selection with a tolerance, where the subset of bands retained



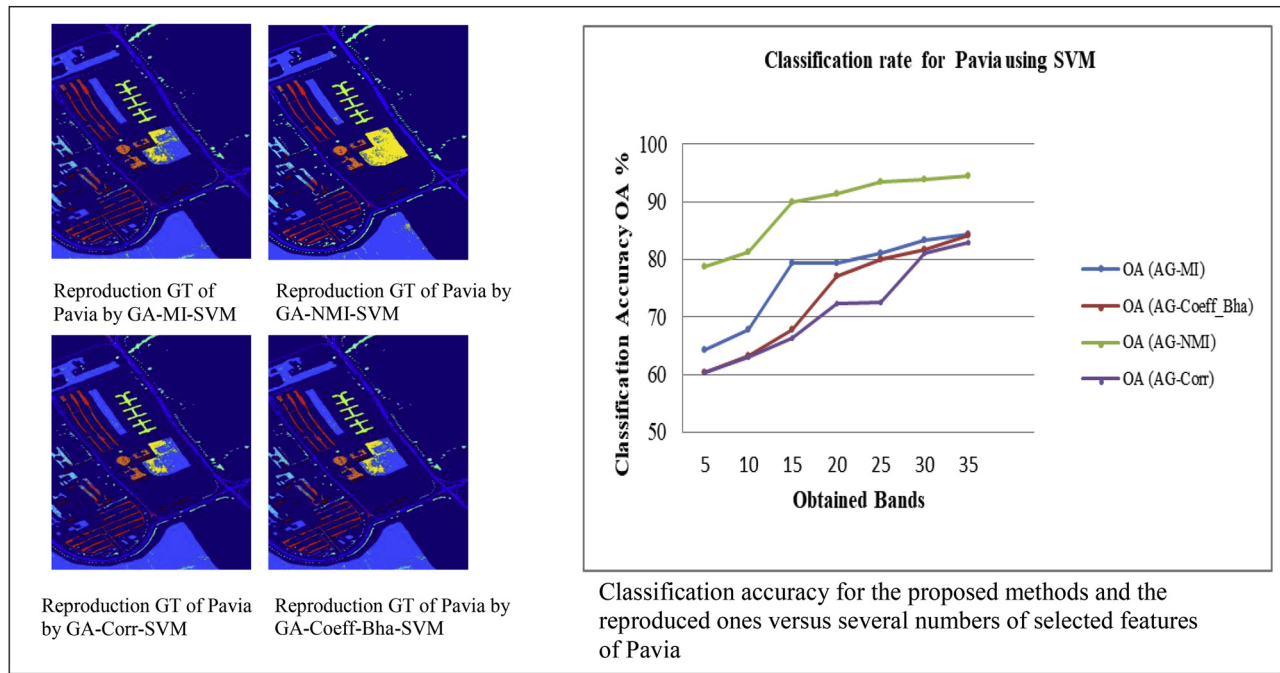
The figures above present a curve that traces the development of the classification accuracy for 80 bands of Indiana Pine, as well as the ground truth reproduced by the use of the selected bands for each method



The figures above show a curve plotting the change in classification accuracy for 80 Salinas bands, as well as the ground truth reproduced by the use of the bands selected for each method

is limited to 82 bands, while the AG-Corr and AG-Coeff-Bha exceed 110 bands in the selection. The 5 bands of SALINAS selected by AG-MI/NMI obtained an OA-IM = 75.24% and OA-NMI = 70.47%. Besides, the overall accuracy with 35 bands selected gets to 94%. Whereas the other fitness functions cannot exceed 66% for 5 bands. Nevertheless, they reach 94% of the OA after a selection of 50 bands.

–The image INDIANA PINE has confirmed what has been synthesized above on the effectiveness of our method in the discrimination of relevant bands. It succeeded in obtaining remarkable values in front of the other reproduced method. The subset of bands selected by the proposed method is limited to 93 bands, while the AG-Corr and AG-Coeff-Bha exceed 150 bands in the selection, which indicates that these methods are not sensitive to the relevant redundant band.



The figures above show a curve plotting the change in classification accuracy for Pavia Dataset, also the ground truth reproduced using the selected bands for each method

Table 1. The classification accuracy for the 80 most discriminating bands retained by the proposed algorithm and compared models for Salinas

Classification results of 80 bands of SALINAS									
Methods		AG-MI		AG-NMI		AG-Corr		AG-Coeff-Bha	
Classifiers		SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN
Classification for each class %	Class 1: Brocoli_green_weeds_1	99.70	99.30	99.90	98.40	99.60	99.30	99.90	99.20
	Class 2: Brocoli_green_weeds_2	100	99.94	100	99.73	100	99.73	100	99.62
	Class 3: Fallow	99.89	99.49	99.69	99.19	99.79	99.29	99.79	99.39
	Class 4: Fallow_rough_plow	99.56	99.42	99.56	99.71	99.71	99.85	98.99	99.42
	Class 5: Fallow_smooth	99.47	98.13	99.77	98.50	99.32	97.68	99.25	97.61
	Class 6: Stubble	100	99.94	99.89	100	99.89	99.79	99.89	99.84
	Class 7: Celery	99.66	99.44	99.72	99.44	99.88	99.44	99.83	99.44
	Class8: Grapes_untrained	90.41	84.24	90.86	83.90	90.62	82.661	89.79	82.94
	Class 9: Soil_vinyard_develop	99.93	99.64	99.90	99.70	99.87	99.12	99.51	97.80
	Class 10: Corn_senesced_green_weeds	98.23	94.32	98.29	94.56	98.10	94.14	97.19	92.12
	Class 11: Lettuce_romaine_4wk	99.81	98.50	99.25	98.31	99.62	94.56	97.19	95.88
	Class 12: Lettuce_romaine_5wk	100	99.79	99.79	99.68	99.68	99.79	99.16	100
	Class 13: Lettuce_romaine_6wk	99.56	98.68	99.56	98.90	99.78	99.56	99.78	98.47
	Class 14: Lettuce_romaine_7wk	98.69	94.95	98.87	95.32	99.25	96.07	98.69	94.01
	Class 15: Vinyard_untrained	82.49	69.29	82.66	68.05	80.46	66.01	82.00	66.67
	Class 16: Vinyard_vertical_trellis	99.77	98.56	99.33	98.89	99.33	98.11	99.77	97.78
OA%		95.42	91.80	95.51	91.57	95.15	90.86	95.01	90.68
AA%		97.95	95.85	97.94	95.77	97.81	95.32	97.54	95.01
KAPPA%		94.90	90.87	95.00	90.61	94.60	89.82	94.45	89.62

5.4. Discussion

In our work, we have delivered a reduction algorithm based on the selection strategy which is founded on the principle of hybridization between GA which has a fitness function that maximizes the criterion function of the integrated filter in

the GA for selection. These criteria function compute the MI and NMI. From the tables and the curves, we deduced the reflected robustness and the stable behavior, also the relevant tolerance in the selection by the proposed method than that reproduced. Hence the max of the bands selected by our method for the three images does not exceed 93 while the

Table 2. The classification accuracy for the 80 most discriminating bands retained by the proposed algorithm and compared models for Indiana pine

Classification results of 80 bands of Indiana pines									
Methods		AG-MI		AG-NMI		AG-Corr		AG-Coeff-Bha	
Classifiers		SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN
Classification for each class %	Class 1: Alfalfa	96.29	77.77	77.77	66.66	88.88	66.66	92.59	66.66
	Class 2: Corn-notill	9,037	74.33	87.30	74.05	81.45	71.54	82.70	74.47
	Class 3: Corn-mintill	84.41	65.22	84.17	70.02	82.73	65.70	81.77	64.26
	Class 4: Corn	83.76	58.11	87.17	49.57	76.06	48.71	81.19	47
	Class 5: Grass-pasture	96.78	93.17	96.78	91.96	95.58	91.16	98.39	93.97
	Class 6: Grass-trees	96.24	97.85	97.31	97.31	96.51	98.39	95.97	96.78
	Class 7: Grass-pasture-mowed	92.30	84.61	92.30	84.61	84.61	84.61	69.23	92.30
	Class 8: Hay-windrowed	98.36	97.95	98.77	97.959	98.36	98.36	98.77	98.36
	Class 9: Oats	100	70	100	50	90	20	100	40
	Class 10: Soybean-notill	85.53	78.71	87.39	82.43	87.80	75.20	81.81	80.78
	Class 11: Soybean-mintill	93.27	81.03	91.57	84.27	89.54	80.79	89.54	80.87
	Class 12: Soybean-clean	91.53	53.42	89.57	55.70	88.59	52.44	87.62	50.16
	Class 13: Wheat	98.11	98.11	99.05	97.16	99.05	100	100	97.16
	Class 14: Woods	97.68	94.12	96.75	95.36	97.99	96.44	95.20	94.74
	Class 15: Buildings-Grass-Trees-Drives	70	53.15	74.21	50	72.10	42.63	75.78	41.57
	Class 16: Stone-Steel-Towers	95.74	82.97	95.74	93.61	85.10	91.48	85.10	93.61
OA%		91.58	80.01	90.93	81.34	89.69	78.83	88.63	79.27
AA%		91.90	78.78	90.99	77.54	88.52	74.01	88.48	75.79
KAPPA%		90.39	77.20	89.66	78.67	88.24	75.81	87.04	76.33

Table 3. The classification accuracy for the 36 most discriminating bands retained by the proposed algorithms and compared models for Pavia

Classification Results of 36 Bands of Pavia									
Methods		AG-MI		AG-NMI		AG-Corr		AG-Coeff-Bha	
Classifiers		SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN
Classification for each class %	Class 1: Asphalt	88.11	85.15	94.20	91.58	87.60	82.32	86.30	84.13
	Class 2: Meadows	98.52	95.49	98.24	98.20	98.31	94.63	98.08	94.42
	Class 3: Gravel	66.82	63.29	80.17	78.45	64.82	58.72	66.82	58.43
	Class 4: Gravel	80.28	76.50	96.80	86.81	75.71	61.22	80.15	66.31
	Class 5: Painted metal sheets	99.55	99.10	99.85	99.40	99.70	98.36	99.55	98.66
	Class 6: Bare Soil	33.00	41.70	89.78	77.13	27.87	33.32	27.83	33.00
	Class 7: Bitumen	80.75	79.54	85.86	88.12	78.04	72.78	79.69	76.39
	Class 8: Self-Blocking Bricks	85.38	80.33	91.09	86.20	85.71	78.65	83.54	75.71
	Class 9: Shadows	99.57	98.94	100	100	98.09	98.30	98.52	98.94
OA%		84.71	83.01	94.72	91.64	83.43	79.50	83.41	79.89
AA%		81.33	80.01	92.89	89.54	79.54	75.37	80.05	76.22
KAPPA%		78.98	76.88	92.98	88.80	77.13	71.85	77.16	72.45

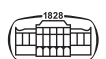
other overcome 150 with an overall accuracy close to that obtained 90 bands by our proposed method. In brief, the important filter combined with the GA in this work is based on the measurement of information when using MI and NMI.

Our method has proved its effectiveness on this type of image; it can reach a high accuracy rate for several images. We had OA = 91.58%, AA = 91.90%, KA = 90.39%, for the 80 bands retained of INDIANA PINE data sets. Otherwise, the proposed method is smooth; it can be adapted to different sizes of images where the result proves this statement. The GA-MI achieved a good rate for the large image

(Salinas image) used in these experiment. 80 bands are selected with OA = 95.76%, AA = 98.06%, KA = 95.28%.

Also for Pavia, although its size is identical to that of other images, we have OA = 94.72 AA = 92.89%, KA = 92.98% for only 36 bands, which distinguishes it from the other image of the classification. As can be seen in the previous table, although our methods are more powerful than the GA-Corr and AG-Coeff-Bha, for all the measures already mentioned.

–For PAVIA, AG-NIM exceeds AG-MI with a large gap of almost 10% for 35 bands.



–It seems that the two functions have the same behavior in the selection of INDIANA PINE. Indeed we can touch a small advantage of the MI compared to NMI with a difference that has approximately 0.60%.

–Lastly, the proposed method didn't disappoint us in front of the largest images SALINAS used in this work. The NMI was also ranked first against the MI with some numbers after the decimal point such as OA = 95.76 by GA-MI and OA = 95.74 by the AG-NMI for the numbers of bands = 82.

6. CONCLUSION

The present work adopted a hybrid method that proved its strong presence among several reproduced methods. The experimental results exposed in the form of tables and curves existing here over, demonstrate the great achievement of the proposed method. It offers a reduced number of bands selected by AG-MI/AG-NMI that reached a high value of OA (+94%). However, the number obtained by these methods AG-Corr and AG-Coeff-Bha doesn't necessarily obtain an optimal subset of bands since it exceeds 120 bands. So they couldn't discern the relevant bands from the informative or non-informative redundant one.

To conclude, the measurement of the quantity of information by MI or NMI often presents higher individual performances in the selection of the bands distinguished by the relation between them. So the association of MI and GAs can get over the Hughes phenomenon. This method was able to ignore non-informative redundancy and choose only the complementary bands which increased the classification rates.

In the future, we will address other fitness functions that will be diversified to handle all the characteristics of the hyperspectral image. Like the used strategies, the kind of spectral or spatial data to process.

REFERENCES

- [1] L. Wang and C. Zhao, *Hyperspectral Image Processing*, Springer, 2016.
- [2] S. Tadjudin and D. Landgrebe, "Classification of high dimensional data with limited training samples," *docs.lib.purdue.edu*, 1998.
- [3] P. K. Varshney and M. K. Arora, *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*, Springer Science & Business Media, 2004.
- [4] K. Koonsanit, C. Jaruskulchai, and A. Eiumnoh, "Band selection for dimension reduction in hyperspectral image using integrated information gain and principal components analysis technique," *Int. J. Machine Learn. Comput.*, vol. 2, no. 3, p. 248, 2012.
- [5] H. Su, B. Yong, P. Du, H. Liu, C. Chen, and K. Liu, "Dynamic classifier selection using spectral-spatial information for hyperspectral image classification," *J. Appl. Remote Sens.*, vol. 8, no. 1, p. 085095, 22 August 2014.
- [6] X. Liu and Y. Bo, "Object-based crop species classification based on the combination of airborne hyperspectral images and LiDAR data," *Remote Sensing*, vol. 7, no. 1, pp. 922–950, 2015.
- [7] J. Ren, T. Kelman, and S. Marshall, "Adaptive clustering of spectral components for band selection in hyperspectral imagery," in *Hyperspectral Imaging Conference*, Glasgow, United Kingdom, 2011, pp. 90–93.
- [8] A. Ghosh, A. Datta, and S. Ghosh, "Self-adaptive differential evolution for feature selection in hyperspectral image data," *Appl. Soft Comput.*, vol. 13, no. 4, pp. 1969–1977, 2013.
- [9] S. Foithong, O. Pinngern, and B. Attachoo, "Feature subset selection wrapper based on mutual information and rough sets," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 574–584, 2012.
- [10] R. Archibald and F. George, "Feature selection and classification of hyperspectral images with support vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 674–677, 2007.
- [11] H. Huang and M. Yang, "Dimensionality reduction of hyperspectral images with sparse discriminant embedding," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5160–5169, 2015.
- [12] L. Ma, M. M. Crawford, and J. Tian, "Anomaly detection for hyperspectral images based on robust locally linear embedding," *J. Infrared Millim. Terahertz Waves*, vol. 31, no. 6, pp. 753–762, 2010.
- [13] B. Arad and O. Ben-Shahar, "Filter selection for hyperspectral estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017.
- [14] M. Merzouqi, E. Sarhrouni, and A. Hammouch, "Classification and reduction of hyperspectral images based on motley method," in *Proceedings of the Computational Methods in Systems and Software*, Cham, Springer, 2019.
- [15] Q. Du, "Modified Fisher's linear discriminant analysis for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 503–507, 2007.
- [16] Q. Gu, Z. Li, and J. Han, "Generalized fisher score for feature selection," *arXiv preprint arXiv:1202.3725*, 2012.
- [17] Y. Bo, W. Ning, and L. Chuanrong, "A band selection method based on synthesized information content for hyperspectral pixel un-mixing," in *IOP Conference Series: Earth and Environmental Science (EES)*, Vol. 17. No. 1, 2014.
- [18] A. Villa, J. Chanussot, C. Jutten, J. A. Benediktsson, and S. Moussaoui, "On the use of ICA for hyperspectral image analysis," in *2009 IEEE International Geoscience and Remote Sensing Symposium*, vol. 4, IEEE, 2009.
- [19] B. Guo, S. R. Gunn, R. I. Damper, et al., "Band selection for hyperspectral image classification using mutual information," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 522–526, 2006.
- [20] A. Martínez-Uso, F. Pla, J. M. Sotoca, et al., "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, 2007.
- [21] M. Merzouqi, H. Nhaila, E. Sarhrouni, and A. Hammouch, "Improved filter algorithm using inequality fano to select bands for HSI classification," in *2015 Intell Systems and Computer Vision (ISCV)*, IEEE, 2015.
- [22] Y.-W. Huang and Z.-S. HE, "A smoothing technique for statistical language model based on mutual information," *J. Chin. Inf. Process.*, vol. 4, 2005.
- [23] J. Feng, L. C. Jiao, X. Zhang, et al., "Hyperspectral band selection based on trivariate mutual information and clonal selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4092–4105, 2013.
- [24] B. Guo, S. Gunn, B. Damper, and J. Nelson, "Adaptive band selection for hyperspectral image fusion using mutual information."



- in 2005 7th International Conference on Information Fusion 2005 July 25, vol. 1, pp. 8 pp.-, IEEE.
- [25] M. Kamandar and H. Ghassemian, "Maximum relevance, minimum redundancy band selection for hyperspectral images," in 2011 19th Iranian Conference on Electrical Engineering, IEEE, 2011.
 - [26] M. Cui, S. Prasad, M. Mahrooghi, L. M. Bruce, and J. Aanstoos, "Genetic algorithms and linear discriminant analysis based dimensionality reduction for remotely sensed image analysis," in 2011 IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2011.
 - [27] C. Harpham, C. W. Dawson, and M. R. Brown, "A review of genetic algorithms applied to training radial basis function networks," *Neural Comput. Appl.*, vol. 13, no. 3, pp. 193–201, 2004.
 - [28] F. Tan, X. Fu, Y. Zhang, et al., "A genetic algorithm-based method for feature subset selection," *Soft Comput.*, vol. 12, no. 2, pp. 111–120, 2008.
 - [29] V. Chahkandi, M. Jalali, M. Mirshahi, et al., "Feature Selection with Hybrid Mutual Information and Genetic Algorithm," in *Proceedings of the International Conference on Data Science (ICDATA)*, 2011.
 - [30] J. Carr, "An introduction to genetic algorithms," *Senior Project*, vol. 1, no. 40, p. 7, 2014.
 - [31] P. Moscato, "On evolution, search, optimization, genetic algorithms and, martial arts: Towards memetic algorithms," *Caltech Concurrent Comput. Program, C3P Report 826*, p. 1989, 1989.
 - [32] Y. Davidor, "Epistasis variance: Suitability of a representation to genetic algorithms," *Complex Syst.*, vol. 4, no. 4, pp. 369–383, 1990.
 - [33] X. Ye, Y. Li, L. Tong, and L. He, "Remote sensing retrieval of suspended solids in Longquan Lake based on GA-SVM model," in 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, 2017.
 - [34] S. Li, H. Wu, D. Wan, et al., "An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine," *Knowledge-Based Syst.*, vol. 24, no. 1, pp. 40–48, 2011.
 - [35] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Machine Learning*, vol. 3, pp. 95–99, 1988.
 - [36] D. Whitley, "A genetic algorithm tutorial," *Stat. Comput.*, vol. 4, no. 2, pp. 65–85, 1994.
 - [37] P. A. Estévez, M. Tesmer, C. A. Perez, et al., "Normalized mutual information feature selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, 2009.
 - [38] E. Gharaati and M. Nasri, "A new band selection method for hyperspectral images based on constrained optimization," 2015 7th Conference on Information and Knowledge Technology (IKT), IEEE, 2015.
 - [39] H. Gao, L. Xu, C. Li, et al., "A new feature selection method for hyperspectral image classification based on simulated annealing genetic algorithm and Choquet fuzzy integral," *Math. Probl. Eng.*, vol. 2013, 2013.
 - [40] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, 2005.
 - [41] F. C. Alcaraz and M. A. Rajabpour, "Universal behavior of the Shannon mutual information of critical quantum chains," *Phys. Rev. Lett.*, vol. 111, no. 1, p. 017201, 2013.
 - [42] M. A. Hossain, X. Jia, and M. Pickering, "Improved feature selection based on a mutual information measure for hyperspectral image classification," in 2012 IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2012.
 - [43] C. Simin, Z. Rongqun, C. Wenling, et al., "Band selection of hyperspectral images based on Bhattacharyya distance," *WSEAS Trans. Inf. Sci. Appl.*, vol. 6, no. 7, pp. 1165–1175, 2009.
 - [44] A. Amine, A. El Akadi, M. Rziza, et al., "Ga-SVM and mutual information-based frequency feature selection for face recognition," *INFOCOMP J. Comput. Sci.*, vol. 8, no. 1, pp. 20–29, 2009.
 - [45] J. Huang, Y. Cai and X. Xu, "A hybrid genetic algorithm for feature selection wrapper based on mutual information," *Pattern Recognition Lett.*, vol. 28, no. 13, pp. 1825–1844, 2007.
 - [46] H. L. Yu, J. M. Sullivan, H. R. Ghadyani, et al., "Image registration by global maximization of mutual information based on a genetic algorithm," *Int. Soc. Magn. Reson. Med.*, vol. 13, p. 1567, 2005.
 - [47] K. Bhardwaj and S. Patra, "An unsupervised technique for optimal feature selection in attribute profiles for spectral-spatial classification of hyperspectral images," *ISPRS J. Photogr. Remote Sens.*, vol. 138, pp. 139–150, 2018.
 - [48] T. A. Warner and M. C. Shank, "Spatial autocorrelation analysis of hyperspectral imagery for feature selection," *Remote Sensing Environ.*, vol. 60, no. 1, pp. 58–70, 1997.
 - [49] G. Liu and F. Zhao, "Efficient compression algorithm for hyperspectral images based on correlation coefficients adaptive 3D zerotree coding," *IET Image Process.*, vol. 2, no. 2, pp. 72–82, 2008.
 - [50] S. V. M. Vishwanathan and M. Narasimha Murty, "SSVM: a simple SVM algorithm," in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, vol. 3, IEEE, 2002.
 - [51] P. R. Marpu, P. Gamba, and I. Niemeyer, "Hyperspectral data classification using an ensemble of class-dependent neural networks," in 2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, IEEE, 2009.
 - [52] D. Manolakis, R. Lockwood, and T. Cooley, "On the spectral correlation structure of hyperspectral imaging data," in *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*, vol. 2, IEEE, 2008.
 - [53] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Science & Business Media, 2009.
 - [54] Y. Guo, S. Han, Y. Zhang, et al., "Guided filter based deep recurrent neural networks for hyperspectral image classification," *Proced. Comput. Sci.*, vol. 129, pp. 219–223, 2018.
 - [55] T. Dutta, S. Dey, and S. Bhattacharyya, "Automatic clustering of hyperspectral images using Qutrit based particle Swarm optimization," in *Intelligence Enabled Research*, Singapore, Springer, 2020, pp. 21–31.
 - [56] A. Weissman, "Mutual information item selection in adaptive classification testing," *Educ. Psychol. Meas.*, vol. 67, no. 1, pp. 41–58, 2007.

