International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V69I11P209

Volume 69 Issue 11, 71-82, November, 2021
©2021 Seventh Sense Research Group®

A Study of A Proposed Suboptimal Selection Strategy Based On Genetic Algorithm And Filters of Mutual Information

Merzouqi Maria[1], Agouzal Mehdi[2], Sarhrouni El kebir[3], Hammouch Ahmed[4]
[1]Electronic Systems, Sensors, and Nano biotechnologies (E2SN), ENSET Mohammed V University Rabat, Morocco
[2]Laboratory of Engineering, Industrial and Management of Innovation (LEIMI), Faculty of Science and Technology, University Hassan 1st, Settat, Morocco
[3] Electronic Systems, Sensors, and Nano biotechnologies (E2SN), ENSET Mohammed V University Rabat, Morocco
[4] Electronic Systems, Sensors, and Nano biotechnologies (E2SN), ENSET Mohammed V University Rabat, Morocco
[1]maria.merzouqi@um5s.net.ma, [2]mehdi.agouzal@gmail.com [3]sahrouni463@yahoo.fr. , [4]hammouch_a@yahoo.com

Abstract - *A novel approach to reduce the dimensionality of hyperspectral images which adopts the selection strategy has been proposed in this work. This approach introduces new genetic algorithm hybridization with one of the mutual information filters (CIM, MRLR, MRMR, NMIFS, MIFS_U2, MIFSU, and MIFS), which have been integrated as its fitness function. Also, this work makes a comparative study of the behavior of genetic algorithms with these filters in the selection of bands. The validation of the result is calculated by two classifiers (RBF-SVM and KNN) to judge the behavior of the hybridization introduced using the overall accuracy (OA %). The new approach is performed on hyperspectral datasets (Pavia, Indian Pines, and Salinas). The results obtained reveal that the performance of this hybridization in selecting the minimum of bands (ex: 30 bands of Salinas) exceeds 95% of OA. According to the results, this approach proves that it can select the most relevant bands and control the redundancy positively. In terms of classification accuracy, they have fascinating results in an acceptable processing time. The results confirm that this hybridization can discriminate between the relevant and irrelevant redundancy bands, efficiently reducing the HSI. Especially the first proposal CMI_GA, except that other comparison hybridization could also fight against the large dimension of these images, precisely those which are not controlled by the adjustable parameter (MRLR_GA with OA =93.93% for only 25 bands of Pavia, and OA=93.44 % for 30 bands of Indiana Pine).*

Keywords - *Reduction of dimensionality, mutual information filters, genetic algorithm, classification, RBF-SVM, hybridization strategy.*

I. INTRODUCTION

The Hyperspectral Image (HSI) is a three-dimensional cube containing spatial information in 2D (XY) and other spectral in 1D (λ), where each of the symbols (X, λ, Y) respectively indicate the spatial height, the number of spectral bands, and the spatial width of the hypercube. Due to the nanoscale resolution, hundreds of the spectral bands taken are highly correlated and closely contiguous over the range (0.4-2.5µm). Referable to the precision and ability to identify negligible differences between objects and their temperature and humidity, HSI can produce the spectral signature of all objects in a scene. Its wealth of information explains its intensive utility in a broad scope of applications. Mineralogy (e.g., for the identification and differentiation between various minerals) [1], also become widespread in the field of defense and the civil situation (ex: border surveillance, drug trafficking, etc.)[2, 3], in agronomy (monitoring and control of crop health)[4]; medicine (ex: identification of residual tumors…) [5].

On the other hand, this image has several problems faced with the good exploit of this pile of data. A hard problem such as (the need for fast computers, more sensitive sensors, great storage capacity…), and soft problem (exceeding the complexity of the computing, ensuring lossless compression…). However, the analysis of these large-dimensional data also faces the problem associated with Bellman's "dimension curse," which is due to the unbalance of the ratio between a large number of learning samples (sample pixels) and a large number of entities (spectral bands) also called Hughes phenomenon [6]. Meanwhile, the multitude of spectral bands poses another overwhelming challenge that arises from the existence of redundant and other non-informative bands that lead to the misunderstanding of the data generation processes, likewise increasing the learning speed of the machine. In addition, all of these factors present obstacles that researchers must confront. Therefore, dimensionality reduction [7] was among the most relevant solutions. It has become a necessary step before classification to avoid confusion of the classifier when predicting, where their advantages are manifested in increasing precision and optimizing the processing time.

A. Reduction by selection

One of the most useful methods to deal with this kind of data is dimensionality reduction [8]. It has become an essential step before applying other methods of analysis or processing this type of image. The process of downsizing is a tricky task, as it consists of scaling down the pile of bands to a set of bands that consists of ten or even twenty bands under the condition of constantly preserving the relevant attributes and even maintaining the helpful redundant while getting rid of other bands that carry no additive information.

This reduction is made according to two possibilities: reduction by extraction or by the selection of attributes. The extraction [15] is done according to several supervised methods such as linear discriminant analysis [8, 9] or PCA [10, 11], and unsupervised by principal component analysis (ICA) [12] and more. It is used on numerous types of hyperspectral images (rural, medical...). It also serves to project the hyperspectral image into a smaller feature space by applying either linear or non-linear transformations [13, 14] to retain the significant components for classification. In addition, the applied transformation distorted the original data. Furthermore, some extraction methods treat each pixel independently [16] without considering the spatial context, which leads to decreased precision during classification.

Hence, feature selection [17] is often used for dimensionality reduction of HSI, as it always preserves the physical significance of data. It doesn't apply any transformation to the initial data. It is founded on the introductory group optimization to find the smaller subset of the necessary bands, which helps describe the target concept [18]. The reductions by selecting the characteristics depend on getting the best subset of bands, which should offer high classification precision and ensure high-class separability. It is applied to eliminate irrelevant, redundant bands and noisy ones.

These reduction methods are based on three fundamental strategies (filter, wrapper, embedded). The filtering strategy is quick [19]. It is based on two essential points. The first is the correct choice of the objective function; the second is the search strategy adopted which must be appropriate to the problem and the type of data [20]. The calculation with the filter is done independently of the classifier. The evaluation of the selected subsets is done by measuring the function, which can be either: gain, distance, score, or consistency [21, 22, 23, 24].

On the contrary, in the Wrapper strategy [25], the classifier is an integral part of the calculation. It uses the classifier as an objective function in finding the best subset. Then the classifier intervenes for each subset of features considered, which explains their excessive computing time. Despite its best results and good performances, this point of time remains a constraint in front of the good exploit of this strategy. For this reason, another strategy has been introduced to minimize and facilitate the computation of the wrapper stage. The hybrid method introduces the filter as an initial step for the wrapper to reduce the first number of features that will be the initial data to be reduced with the wrapper [26, 27, 28]. As an add-on, the embedded approach takes advantage of the benefit of the two approaches referred above. The virtue of this approach with respect to others is presented in Classifier learning, which is executed in parallel with the selection [29].

First, this work proposes hybridization for selecting attributes based on a genetic algorithm (in short, FSGA), which adopts the principle of eliminating unnecessary, redundant attributes and insignificant attributes from the initial data while preserving all the relevant attributes without destroying the original form of data. FSGA can benefit from the robustness and potential of genetic algorithm selection, which will take the strain off the classifier in several ways. Most important for us is strengthening the prediction while avoiding the "curse of dimensionality" due to the induced knowledge provided by the selection step and the decrease of the computation time. Not only this, this work aims to study the behavior of the proposed genetic algorithm with selection filters based on mutual information (in short, MIFs), which are presented in this procedure as the fitness function of the used genetic algorithm. Each of the methods has its advantages. The motivation behind this comparative study is to evaluate the efficient filter with genetic algorithms in the reduction of dimensionality of HSI.

This work is organized as follows. In the sect. 2, a theoretical part that reminds all the filters used and a brief presentation of the principle of genetic algorithms. Then the sect. 3 is occupied for the explanation of the proposed method and submit the tools used. Sect. 4 will be devoted to presenting the results, with a comparative analysis between the different algorithms. Sect. 5 will be dedicated to discussing the general synthesis, which allows us to record the evolution of the different selection methods. In the sect. 6, presents the conclusions of this work.

II. LITERATURE REVIEW OF FILTERS (MIFs) AND GA

The used selection strategy is founded on a genetic algorithm whose fitness function is one of the filters based on mutual information [30].

The choice of mutual information [31] filters always stems from the success of this family. It's recognized by its ability to deal with all kinds of problems in several fields. It is used to measure statistical dependence. Two substantial properties distinguish it in front of other indicators: The ability to quantify the dependence between random variables, regardless of their relationship (linear or nonlinear); stability and immunity against the spatial transformation of characteristics that can be differentiable or even invertible. So this study focuses on feature selection filters using mutual information as an objective function. Fano [32] proved that if there are a significant likelihood and a quantity of information between the transformation data (candidate bands, noted by: f_c) and the desired target (ground truth noted by C), that implies a lower bound of the error probability. The delimitation below presents the Fano inequality [33]:

$$\frac{H(C)-I(f_c;C)-1}{\log N_c} \le P_e \le \frac{H(C)-I(f_c;C)}{\log 2}. \qquad (1)$$

With H (C) is the entropy of the output. In this case, the output is the ground truth C; it also $I(f_c;C)$ represents the mutual information between the candidate band f_c , and the ground truth C. N_c represents the number of output classes.

This idea inspires Battisti's MIFS filter concept [34, 35]. This filter permeates almost all scientific work, such as references or even analytical methods, especially those dealing with selecting attributes. Battiti's selection process selects the chosen number of most relevant attributes (k) from the original set of attributes (F). Adopting the principle of incrementation by attributes to fight against the combinatorial explosion of all possible subsets of attributes belonging to the original set (F). Generally, filters are based on two pillars: an objective function and a research strategy. Battiti's filter is based roughly on four hypotheses: The proposes as an initial step the classification of all the attributes; to differentiate between relevance and redundancy introduced a heuristic selection function, respecting a greedy search strategy to have a subset of characteristics that must be relevant and optimal.

A. The MIFS algorithm

The algorithm process can be formulated as follows:

1 Process initialization:

- The initial data set: is | F |, which has the input characteristics
- Definition of an empty set | S |
- Determination of the desired number of characteristics k
- Determination β

2 Calculation of mutual information between each variable with the output class:

$\forall f_C \in F$, calculate $I(f_C;C)$

3 Choice of the first characteristic:

- The chosen characteristic must have a high value of mutual information:

Max $I(f_C;C)$

- Reset F ← F \ { f_C }, define S { f_C }

4 Greedy selections:

- Repeat step 3 until | S | = k

A Calculation of MI between all pairs of entities:

- Calculate $I(f_C; f_S)$ with $\forall f_C \in F$ and $f_S \in S$ If it's not already available

B Selection of the following characteristic:

The selection of the character if it maximizes the following function

$$\forall f_C \in F \quad \max\left[I(f_C;C) - \beta \sum_{f_C \in F} I(f_C; f_S) \right]. \qquad (2)$$

- Reset F ← F \ { f_C }, define S { f_C }

5 Exit the program when the k reaches:

- The set S contains the selected characteristics.

$I(f_C; f_S)$ Represent the measure of redundant information between the input candidate bands f_C and those already selected f_S. Hence β is an important parameter in this formula; it takes a value that varies from 0 to 1. The identification of this value depends on the type of data used to fix it empirically. The parameter β exploited to define the relevant redundancy between the candidate band and the set already selected.

The value of β plays a significant role in the efficiency of this filter; if β=0, this overwrites the second part of the filter, making it as a calculation of the mutual information gain between the candidate band and the C classes to select independent bands, while a value of $\beta > 0$ controls the penalty of inter-band redundancy.

.

B. The MIFSU filter

Among the limitations of this filter is that it cannot be said about the adequate value β, which must be standardized regardless of the type of data to be processed. Kwak and Choi [36] proposed an improved filter from MIIFS under the name MIFS-U, but they always kept the setting β. The MIFS-U is dedicated to being more suited to problems where the distribution of information is uniform in space.

$$I(f_C;C) - \beta \sum_{f_S \in S} \frac{I(f_S;C)}{H(f_S)} I(f_S; f_C). \qquad (3)$$

We always notice the existence β in the second term of the formula (3), which is used to estimate the redundant information between the candidate band f_C and the others already selected f_S with respect to the classes C. As already mentioned, the variability of β influence directly on the selection of bands, but the weight of $\frac{I(f_S;C)}{H(f_S)}$ has mitigated the effect of β to some extent. However, this filter also has limitations like the MIFS: choosing the appropriate value β remains unresolved. In addition, if the distributions of the input feature were very distinct from the uniform causes an estimation error.

C. The MIFS_U2 filter

In 2006 and .al [37], Huang improved the MIFS and MIFS_U filters to attenuate the influence of the parameter β acting on the second part of the filter, adding the expression $\frac{1+\beta \log_2^{|S|}}{|S|}$ to the formula of the MIFSU filter.

$$I(C; f_c|S) = I(C; f_c) - \frac{1+\beta \log_2^{|S|}}{|S|} \sum_{f_s \in S} \frac{I(C; f_s)}{H(f_s)} I(f_s; f_c). \qquad (4)$$

S represents the number of bands selected. To estimate $I(C; f_c|S)$, they took the average redundancy f_c for all the entities of S. In addition, and they found from the data experiments that it is important to increase in parallel the second term of the formula (4) and the number of bands selected in S, except that the information increment is reduced. For this objective, they used the logarithmic function to adjust the probable aptitude of variation of this term. Keeping always the parameter β. This will have a fixed value in this study in 0.5.

D. *The NMIFS filter*

Estevez et al. in 2009 [38] use the means of the normalized mutual information (NMI) in its filter with the genetic algorithms to measure the redundancy between the candidate attributes and the subset of the selected attributes. They made the ratio of the NMI by the minimum entropy of the two attributes (candidate, already selected) [39].

$$I(f_c;C)-\frac{1}{|S|}\sum_{f_s\in S}\frac{I(f_s;f_c)}{\min(H(f_c),H(f_s))}. \qquad (5)$$

This NMIFS filter solves the problem of determining the parameter value, which is considered a limitation of the filters above, eliminating it clearly.

E. *The mRMR filter (min-redundancy max-relevance)*

. The NMIFS algorithm only ameliorates the mRMR filter, which was introduced by (Peng et al. 2005) to solve the problem of redundancy (relevant or irrelevant) between the selected attributes. Its principle resides in the calculation of the information, which is usually used to determine the independence of a given random variable with respect to another known one [40].

Calculation of the relevant information:

$$\max P(S;C)=\frac{1}{|S|}\sum_{f_s\in S}I(f_s;C). \qquad (6)$$

Calculation of redundant information:

$$\min R(S)=\frac{1}{|S|^2}\sum_{f_s\in S,\, f_c\in F}I(f_s;f_c). \qquad (7)$$

mRMR serves to minimize redundancy and maximize relevance simultaneously:

Max (relevant; redundant) = Relevant (S; C) - Redundant (S):

$$mRMR=P(S;C)-R(S). \qquad (8)$$

F. *The MRLR filter (most relevant and least redundant)*

MRLR has the same mRMR filter principle. It serves to measure relevance and the elimination of uninformative redundancy among selected attributes. MRLR also combines the Relevant (S; C) "P" and the Redundant (S) "R" in a single O formula. Calculating the same equations (6) and (7) of the mRMR filter [41].

It is used to optimize R and P to simultaneously obtain the most relevant and least redundant attributes, respecting the principle determined below:

$$Maximize\,O(P,R)=\frac{P}{R}. \qquad (9)$$

G. *The CMI filter (conditional mutual information)*

To avoid the drawbacks of the other selection filters already mentioned, Sombut Fishing et al. have proposed another algorithm based on the same principle of mRMR and MRLR. Their filter is based on removing redundancy. Neither between classes, nor their redundancy with respect to another determined attribute, as they can be very similar. It selects relevance using conditional mutual information metrics while avoiding the parameterized β , which greatly impacts the selection decision-making [42].

$$CMI=I(C;f_c)-\frac{H(f_c/C)}{H(f_c)}R(f_c;f_s). \qquad (10)$$

With

$$R(f_c;f_s)=\sum_{f_s\in S}\frac{I(f_c;f_s).I(C;f_s)}{H(f_s)H(C)}. \qquad (11)$$

Formula (11) has almost the same operation as the redundancy calculation term

according to a characteristic already selected in the selection criterion of the MIFS-U filter.

The term $\frac{H(f_c/C)}{H(f_c)}$

is the distinction of this filter from other filters.

This expression was added to measure the relevance of the selected attribute based on the output classes. The reduction of this expression implies that the relation between the next selected attribute (band f_c) and the output classes "C" or the ground truth is important. In other words, this indicates they are strongly related, which explains the increase of the mutual information value between the selected band and the ground truth. Then selection is said relevant. On the other hand, its increase implies a decrease in the mutual information $I(C;f_c)$, which abolishes the f_c band and makes it useless. It can be deduced that this expression has replaced the difficult parameter to assign and is not practical for real problems.

With $H(f_c/C)$ is the conditional entropy, which can be calculated from mutual information as follows:

$$I(f_c;C)=H(f_c)-(f_c/C). \qquad (12)$$

With $H(f_c)$ the Shannon entropy of f_c :

$$H(X)=-\sum_{x\in X}P(x)\log_2 P(x). \qquad (13)$$

With the P(x) is the probability of X, which represents f_c in this case.

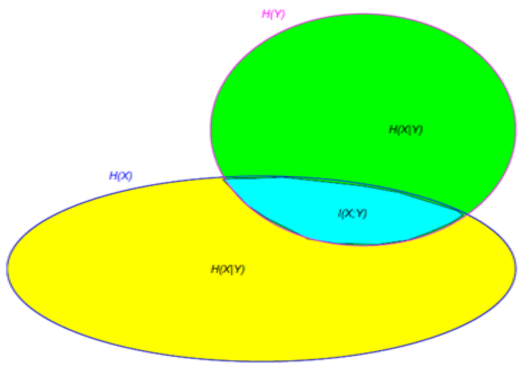


Fig.1. The Venn diagram illustrates the Relations between joint and conditional entropy and mutual information.

These relations can be summarized from the diagram of Venn with X = f_c and Y = C are two variables jointly distribute

These filters were introduced as evaluation metrics (fitness function) for the selection of discriminating subsets of the used genetic algorithm; these subsets will be classified at the final step with the main classifier SVM and recalculated by the KNN classifier to confirm what will be retained as synthesis according to the values obtained from OA by RBF-SVM.

H. Genetic Algorithm

According to the original thoughts of Darwin, who sees that nature and life in this world, with their various surprising forms, and their evolutionary existence, go through processes of mutation and natural selection controllable by the survival capacity of the best and the strongest. Their reflection obtained several assents. It resulted in the temptation to use the principle of "selection and adaptation" of nature and adapt them to resolve different kinds of optimization problems. John Holland is one of the many scholars who believe in Darwin. In 1970, John Holland invented genetic algorithms as a heuristic method. In 1992 he improved it and described it well in his book "Adaptation in natural and artificial systems.” [43].

GA is a mathematical-like iterative algorithm used to find an optimal solution under a fixed parameter with interest. All the individuals in these algorithms go through several “selection, crossing, and mutation” processes, from which they must prove their existence at each iteration in an evolutionary way. In other words, over time, the individual struggles to survive each generation where the most suitable set and which meets the defined parameter appears as the optimal solution. These algorithms have become an efficient and robust tool for solving problems regardless of their size and type: simple or complex. It also has an additional point compared to other algorithms, consisting of the presence and absence of information. This has a significant impact on the bottom line. In addition, it also has a controllable history repository to preserve the information sought not to lose it in other generations. This option is known as elitism, which also sets it apart [30, 44].

They are judged by their fitness function, which is considered to be its core. It is conceived as a real challenge that validates the credibility of the algorithm in solving the problem. This function must therefore be appropriate to the genotype. It is mainly used to calculate a score, gain, distance, or consistency. For the results to be effective and satisfactory, the programmer must carefully introduce the appropriate combination that meets the need exactly.

Usually, genetic algorithms take a user-defined population formed from a set of chromosomes or genotype that contains a set of genes presented by a vector of [0 1] in the case of binary genetic algorithm code, which is generated at random way. 1 indicates the presence of the gene and 0 their absence. The initial population went through the already cited options, which are well described in our previous work [30]. For this reason, these details are not repeated here. These options cause the appearance of other chromosomes, which form a new population, which will be used in future algorithm iterations. The chromosomes will be evaluated at each generation by the fitness function, which represents the dynamo of the algorithm.

As already advised, the fitness function must be built around the characteristic of the problem for it to be successful. In addition, to validate their reliability, also ensure that the selection is credible and optimal. The stopping conditions of a genetic algorithm are generally classified into two choices, either reaching a maximum number of generations or reaching predefined fitness values for the generation.

III. METHOD AND MATERIALS

In this work, the principle of a binary genetic algorithm is adopted. The proposed algorithm starts with an initial population that consists of 40 chromosomes in the first generation. Hence, each chromosome is represented by a vector of length equal to the number of bands in each hyperspectral image used. Each number of positions takes a value [0 or 1] distributed randomly and which respectively determines the absence and the presence of the band. The population will automatically go through the options of the basic genetic algorithm (selection used: roulette wheel; Crossover, with a probability of 0.98, and the mutation with a probability of 0.005) to create a new population at each generation. The number of generations in this work is fixed at 5. These values are taken empirically, considering the quality of the result and the algorithm execution time, which must be reduced. The fitness function (Fit_F) used in this work is one of the filters mentioned above.

The Fit_F will be calculated for each chromosome in each generation to find the optimal chromosomes, which will be summoned in generations until the set number of generations is reached, due to finding the best and strongest chromosome that could have survived during these generations with objectivity. Figure 2 below summarizes what is described here. To evaluate the efficiency of chromosomal selection that obtains a subset of the selected bands. The main RBF-SVM will be used to classify the subset. Additionally, another classifier, KNN-K =5, will be used to repeat the same calculation done to highlight the final confirmation and judgments on the proposed method.

A. The RBF-SVM and KNN classifier

- SVM is among the most accurate and efficient classifiers for managing the classification of hyperspectral images. The experiments launched by our team [45] confirmed that the RBF kernel is the most suitable for validating the results. This is the case studied, which is why SVM with the RBF kernel was adopted. It has proven its robustness in classifying hyperspectral images since it ensures intra-class separation by maximizing the margin M which determines the distance between image classes. For the classifier's training, 50% of each hyperspectral image's selected and random pixels were used to build the training examples, and the others were kept for the tests [30].
- The K-Nearest Neighbor (KNN) classifier is a versatile learning algorithm. Often used for supervised classification (the studied case). Non-parametric and easy to understand, this makes it more useful. It is based on the

similarity measure. The number knows the KNN classifier of neighbors K which must be chosen to control the prediction model. In the studied case, the K = 5.

The classifiers calculate several parameters such as OA (Overall precision), which will be the only one used in this work. Others like AA and KAPPA are also already defined in our previous work. The OA will determine the results of this work to judge the behavior of the proposed genetic algorithm with mutual information filters and their impact on the selection of hyperspectral images. The OA is a metric calculated by the classifier according to the formula (14) below. It is the ratio of correctly classified labels to the samples of global classes. The TP,
TN, FP, FN respectively represent the predictions: true positives, true negatives, false positives, and false negatives [30].

$$OA = \frac{TP + TN}{TP + TN + FP + FN}. \qquad (14)$$

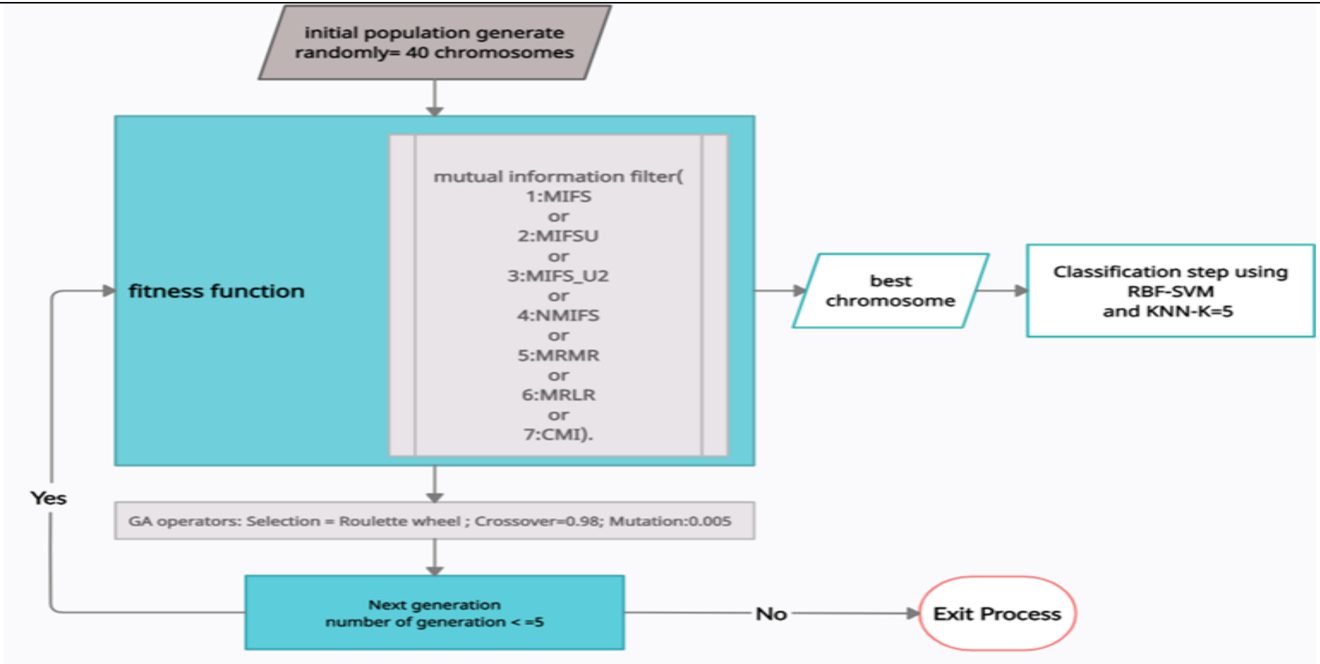


Fig.2. The flowchart process of selection is based on a genetic algorithm using the mutual information filter (MIFs_GA).

As already mentioned, the applied experiments were performed on the data sets from Pavia, Indiana Pine, and Salinas, containing 9, 16, and 16 classes for 103, 220, and 224 reflectance bands as characteristics of the classification process. Each image is characterized by its nature. This explains the choice of the diversity of data in this work.
Pavia is characterized by its urban nature, as it covers a space of low density, with the existence of a set of forms [30]. In contrast, Indiana Pine is subdivided into two areas. The first part occupies 1 third of the image and covers the urban area, which shows: low-density housing, highways, railways, etc.; however, the rest of the image reveals the rural area, which consists of a variety of perennials.
Moreover, Salinas only covers a purely rural territory, but its issue is distinguished by mixed classes, which makes it complex in the treatment. The table below shows more details on each image [30].

TABLE I. OTHER DETAILS ON THE DATASET USED

Data set	**Pavia**	**Indiana Pine**	**Salinas**
Sensor	ROSIS-03	AVIRIS de la NASA	AVIRIS
Place	The University of Pavia, northern Italy	Indiana Northwestern USA	Salinas Valley, California, USA
Frequency band	0,43 to 0,86 µm	0,4 to 2,5 µm	0,4 to 2,5 µm
Band Resolution	1.3m/pixel	20 m/pixel	3,7 m/pixel
Number of bands	Global 115; 103 Used	Global 224 ; 220 Used	224
Number of labeled pixels	42776	10366	54129

IV. RESULTS AND COMPARATIVE ANALYSIS

The following section will be devoted to presenting the results of the hybrid algorithm (MIFs_GA). As already well described, this algorithm is based on the genetic algorithm and one of the mutual information filters calculated in the fitness function to increase the precision of the selection of the hyperspectral images to reduce its dimensionality in a short time. This study is done in a first-place to study the behavior of genetic algorithms with selection filters based on mutual information. It's also a perspective of our work [30], as shown on the curves below(fig 3;4;5) and the tables (table 2;3;4) that illustrate in detail the classification results for each image. This hybridization has proved its success with the CMI filter initially proposed. The MRLR and others have been added to make the comparison; like the mRMR which has also proved their existence by their high value of OA with OA a minimum of the bands. On the other hand, the other filters, especially those condemned by the parameter, could not exceed the forecasts even if their results were satisfactory.

A. Results

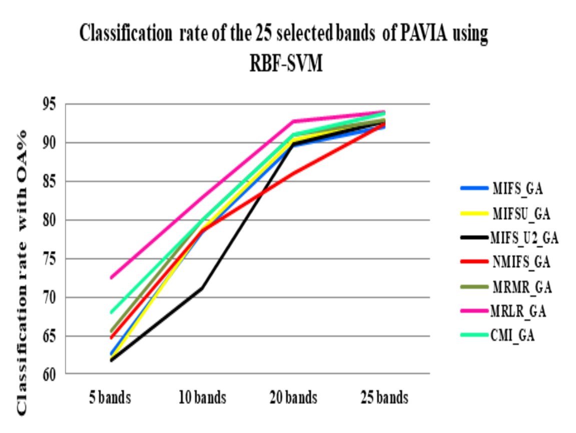


Fig.3. The following curve plots the development classification rate from 5 to 25 bands selected by the proposed hybridization MIFs_GA of PAVIA

The curves (Figure 3, 4, 5) below have been generated to plot the statistical results along two vertical axes, which presents the percentage of the evaluation metric chosen OA as a function of the horizontal axis, which gives the number of bands selected. The proposed method was able to surpass the pre-judgment with its results which demonstrated their robustness. In addition, the results presented in the tables ( )prove that the proposed hybridization overwhelmed with a large positive deviation the results obtained in our previous work [30], which were based on the implementation of mutual information and normalized mutual information, which also had an important result. For 30 Salinas bands using GA_ MI, the OA = 93.60% [30], and the CMI_GA OA = 95.60%.

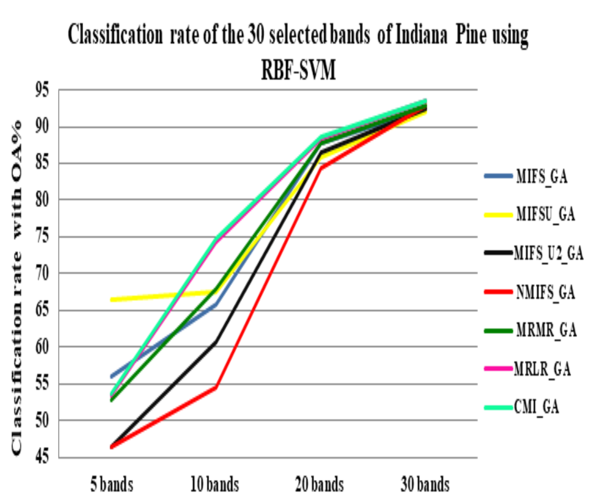


Fig.4. The following curve plots the development classification rate from 5 to 30 bands selected by the proposed hybridization MIFs_GA of Indiana Pine.

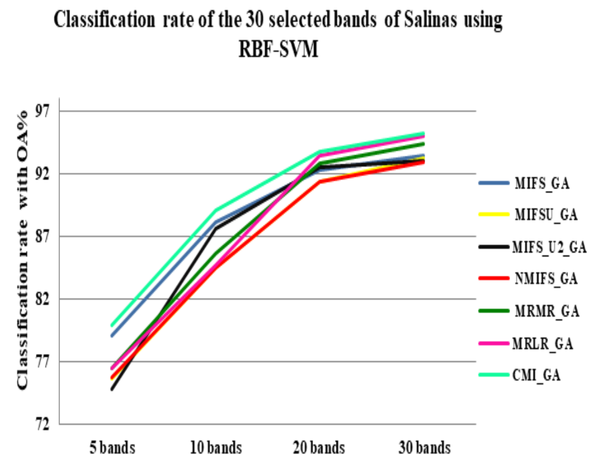


Fig.5. The following curve plots the development classification rate from 5 to 30 bands selected by the proposed hybridization MIFs_GA of Salinas

We deduce as initial analysis and synthesis, as it can be seen that the curves of the three images generally have the same paces, which tend towards growth with a flexible divergence. These results explain the stability and the feasibility of exploiting this hybridization to reduce the dimensionality of hyperspectral images. Moreover, the important values of OA% (93.93; 93.63; 95.24) for (25; 30; 30) of Pavia, Indiana, Salinas respectively proved the efficiency of the introduced algorithm.

TABLE II. THE CLASSIFICATION RATE % FOR THE RETAINED BANDS OF PAVIA FOR ALL THE ALGORITHMS USED IN THIS STUDY USING THE (SVM / KNN)

Number of bands and the used classifier	CLASSIFICATION RATE BY OA % for PAVIA							
	5 bands by SVM	5 bands by KNN	10 bands by SVM	10 bands by KNN	20 bands by SVM	20 bands by KNN	25 bands by SVM	25 bands by KNN
MIFS_GA	62,72	58,61	78,42	73,45	89,67	82,7	91,97	84,95
MIFSU_GA	61,97	57,65	78,79	73,63	90,24	86,39	92,95	84,38
MIFS_U2_GA	61,83	57,16	71,06	65,04	89,87	83,28	92,65	85,89
NMIFS_GA	64,69	60,4	78,57	75,45	86,02	84,85	92,37	85,98
MRMR_GA	65,58	61,12	79,96	76,83	90,92	86,9	92,98	88,38
MRLR_GA	72,46	68,44	82,81	81,34	92,7	88,67	93,93	90,41
CMI_GA	67,98	61,46	79,98	76,95	90,94	86,98	93,71	88,46

TABLE III. THE CLASSIFICATION RATE % FOR THE RETAINED BANDS OF INDIANA PINE FOR ALL THE ALGORITHMS USED IN THIS STUDY USING THE (SVM / KNN)

Number of bands and the used classifier	CLASSIFICATION RATE BY OA % for INDIANA Pine							
	5 bands by SVM	5 bands by KNN	10 bands by SVM	10 bands by KNN	20 bands by SVM	20 bands by KNN	30 bands by SVM	30 bands by KNN
MIFS_GA	55,97	47,96	65,73	52,71	87,75	65,74	92,36	67,57
MIFSU_GA	66,48	48,72	67,62	53,37	85,91	70,16	91,94	68,59
MIFS_U2_GA	46,51	33,61	60,81	47,6	86,47	62,32	92,4	66,32
NMIFS_GA	46,45	33,65	54,59	42,65	84,32	62,77	92,78	65,3
MRMR_GA	52,88	35,73	67,99	53,46	87,86	71,67	92,81	74,83
MRLR_GA	53,52	47,36	74,39	65,8	88,39	79,5	93,44	81,55
CMI_GA	53,69	47,85	74,68	67,15	88,52	79,72	93,63	81,76

TABLE IV. THE CLASSIFICATION RATE % FOR THE RETAINED BANDS OF SALINAS FOR ALL THE ALGORITHMS USED IN THIS STUDY USING THE (SVM / KNN)

Number of bands and the used classifier	CLASSIFICATION RATE BY OA % for SALINAS							
	5 bands by SVM	5 bands by KNN	10 bands by SVM	10 bands by KNN	20 bands by SVM	20 bands by KNN	30 bands by SVM	30 bands by KNN
MIFS_GA	79,13	75,94	88,14	85	92,29	88,08	93,46	88,74
MIFSU_GA	75,68	67,28	84,51	79,6	91,44	87,35	93,28	88,42
MIFS_U2_GA	74,82	72,26	87,67	83,75	92,56	87,78	93,05	88,21
NMIFS_GA	75,83	72,65	84,5	81,55	91,38	87,1	92,97	87,99
MRMR_GA	76,5	75,09	85,65	84,21	92,83	89,97	94,43	90,83
MRLR_GA	76,48	70,15	84,72	82,35	93,43	90,92	95,04	91,8
CMI_GA	79,96	76,07	89,1	85,14	93,81	90,98	95,24	91,95

B. Comparative analysis

In this study, two classifiers have been used (primary RBF-SVM, KNN-K = 5 secondaries) to confirm the reliability of the results of the proposed hybridization.

Firstly, this work was conceived on the proposal of a single method based on CMI and GA as a hybrid strategy for selecting bands of hyperspectral images, which was able to overcome Hugh's phenomenon with success for all three images. As observed, it reached a percentage of OA = 93.81 with the SVM and more than 90% with the KNN for just 20 bands of Salinas, which is known for its complex nature. In this regard, another filter based on mutual information was added to make the comparative study of the behavior of genetic algorithms with this hybridization in the reduction of dimensionality of HSI by the selection approach.

The first place was vacant by the CMI_GA hybridization. It gives important results, indicating that it could overcome the limitations of the other hybridization introduced. It is inferred that it can detect the combinatorial relations between the relevant bands, as well as the appropriate redundancy to some extent, especially in the selection of (10, 20, and 30) bands for Indiana and Salinas, presented respectively by the (OA with SVM= 74.68%; OA=88.52%; OA=93.63%); (OA with SVM=89.1%; OA=93.81%; OA=95.24%).

We find the presence of CMI_GA much less obvious than MRLR_GA for Pavia, which is known for its low attendance of redundant bands. MRLR_GA is ranked first according to these OA values for all measures (from 5 to 25 bands) with both classifiers. For the minimum bands, the OA % with SVM exceeded 72, and KNN = 68.44. It kept its peak up to 25 bands with an OA% greater than 93 with SVM and KNN = 90.41%. This hybridization has fewer limitations since it involves selecting the relevant bands and eliminating the redundant ones according to the ratio used in the formula, which acts on the balance between the redundancy and the relevant.

In addition, the mRMR_GA is classified in the third order after the MRLR_GA and the CMI_GA for the discrimination of the relevancy and the redundant bands. It also has values that cannot be underestimated, namely the percentage of OA reached for the selection of 20 bands from Salinas exceeded 92%, and for Indiana, OA = 87.86%, and for Pavia, OA = 90, 92%.

We cannot deny that the other hybridization did not also have impressive results. It is noted that the MIFSU_GA was able to exceed all the hybridization introduced in the selection of the minimums of the bands (5 bands) for Indiana, which is considered the most correlated image in these three used images, and which contains a lot of redundancy. She had a percentage higher than 66, from where the other hybridization could not reach even 60% of OA. But she did not maintain this progression. This comes down to the redundancy tolerance by a parameter β that must be programmable for each number of bands selected and each kind of data (each image).

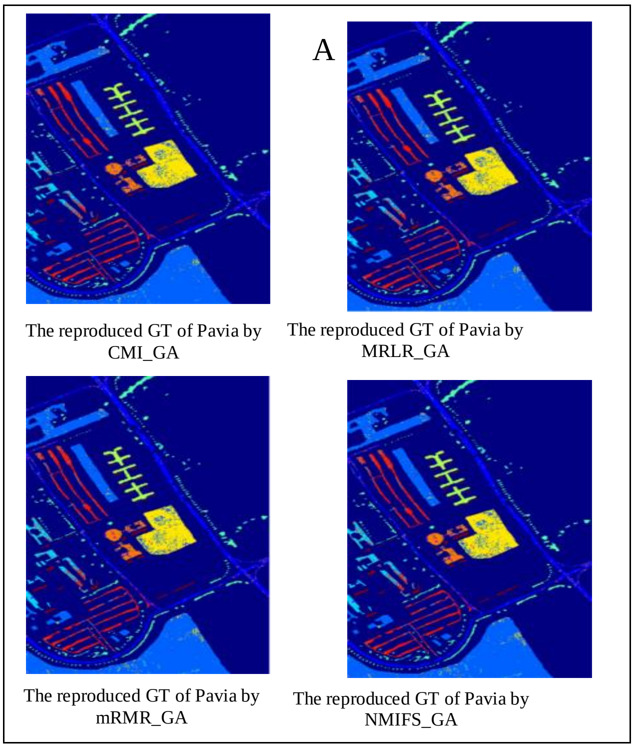
C. Discussion and synthesis

The results deduce that this hybridization is efficient and applicable for reducing hyperspectral images, especially the CMI_GA and MRLR_GA, which reduced the three images with a classification rate expressed by bold values of OA. The mRMR_GA also has a good outcome. It takes the third order with its important results in the selection. This comes down to the balance in these formulas between the redundant and the relevant, except that some hybridization has failed in the mission entrusted to them.

The hybridization MIFS_GA, MIFSU_GA, MIFS_2U_GA have more or less efficient than the others. That constantly comes up in the existence of the parameter β , which was set at 0.5 for the three images. Also, the MIFS filter with the GA marked its presence in a single image for a minimum of the bands (5, 10), which entails the absence of redundancies. These results confirmed that the value choice β should be varied according to the data and even according to the number of the selected bands. In addition, if the term devoted to calculating the redundancy between the selected band and the next selected one becomes negligible compared to the term on the right, the selection algorithm will be forced to select non-redundant bands with those already selected. However, this will cause irrelevant bands to be selected earlier than the relevant redundant. So the use of these filters requires human intervention in this selective operation. Above all, this will cause heavy selection. The hybridization is based on the genetic algorithm and the NMIFS, which is considered an improvement of the previous three filters. The NMIFS tried to balance redundancy and relevance by eliminating the manual parameter β as noticed that the two terms of the subtraction in the formula are in the range [0, 1]. If the number of bands increases, this will cause the neglect of the second term, which controls the redundant penalty [38]. In this case, the mutual information on the right hand will be dominant. In addition, NMIFS generally suffers from the same limitation as that of the three filters mentioned above when β it is too small. Also, the NMIFS

According to its formula, the term which controls the redundancy can take different weights of normalization of the bands. This explains its fall as an appropriate function of the genetic algorithm. Furthermore, this failure is represented in the poor results of its selection, as already shown in the tables above.

To this end, the final synthesis is that these three configurable filters and the NMIFS are not applicable as a fitness function for the genetic algorithms in selecting hyperspectral images.

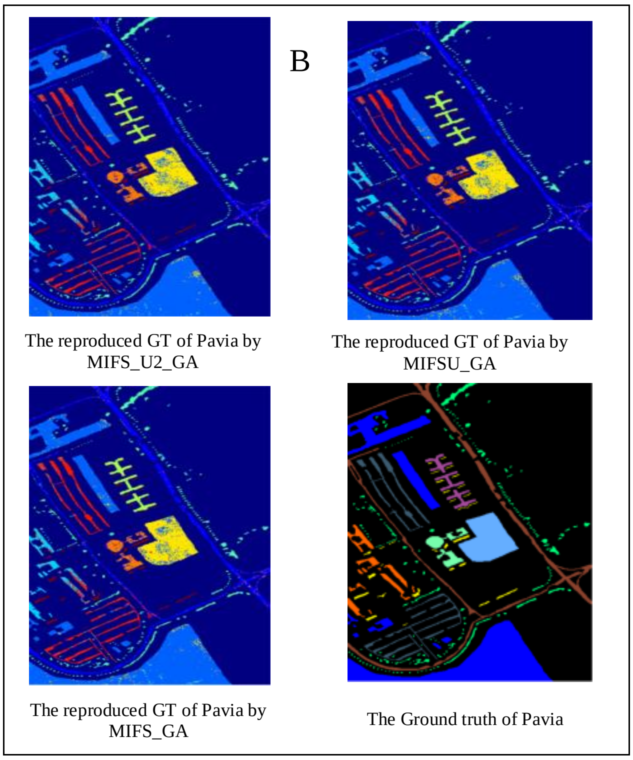


The reproduced GT of Pavia by CMI_GA

The reproduced GT of Pavia by MRLR_GA

The reproduced GT of Pavia by mRMR_GA

The reproduced GT of Pavia by NMIFS_GA



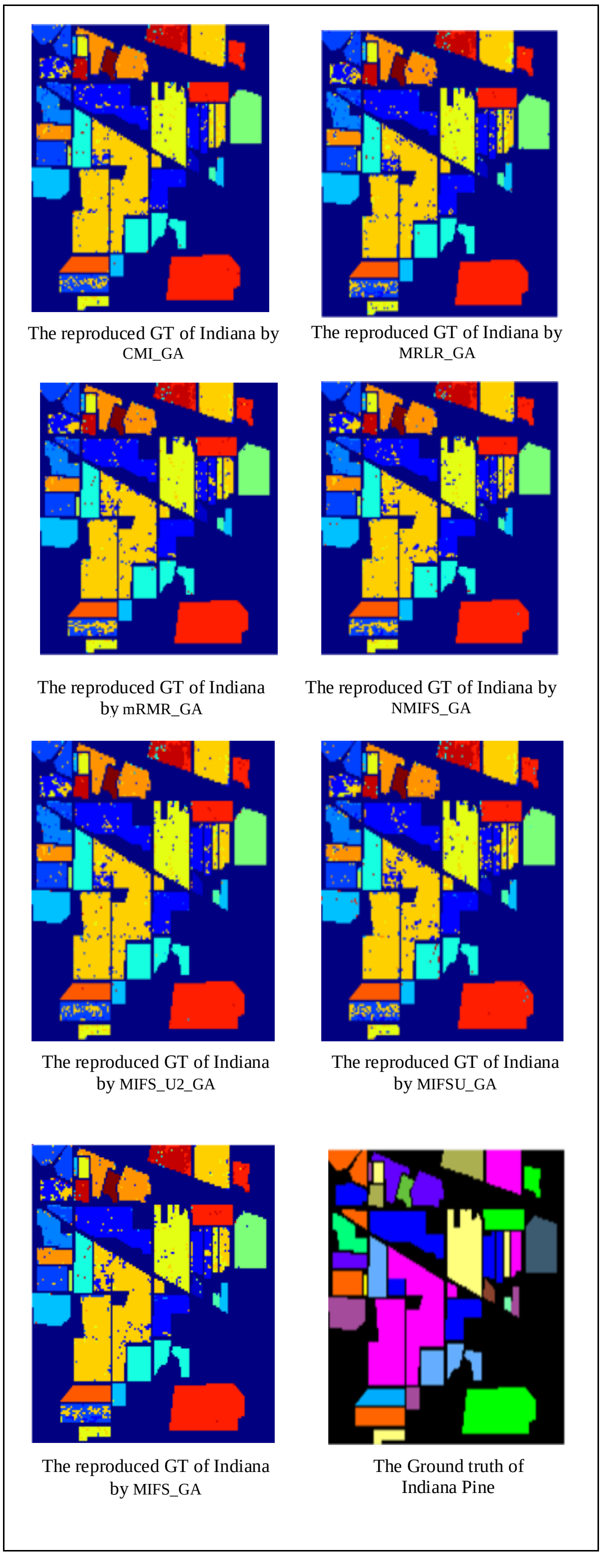
The reproduced GT of Pavia by MIFS_U2_GA

The reproduced GT of Pavia by MIFSU_GA

The reproduced GT of Pavia by MIFS_GA

The Ground truth of Pavia

Fig.6. (A, B) presented The reproduced (GT) by the proposed hybridizations for 25 bands of Pavia and the reel GT.



The reproduced GT of Indiana by CMI_GA

The reproduced GT of Indiana by MRLR_GA

The reproduced GT of Indiana by mRMR_GA

The reproduced GT of Indiana by NMIFS_GA

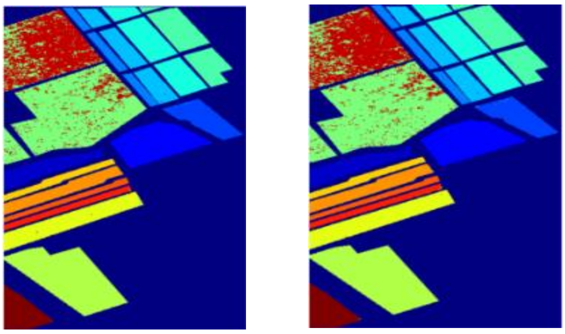
The reproduced GT of Indiana by MIFS_U2_GA

The reproduced GT of Indiana by MIFSU_GA

The reproduced GT of Indiana by MIFS_GA

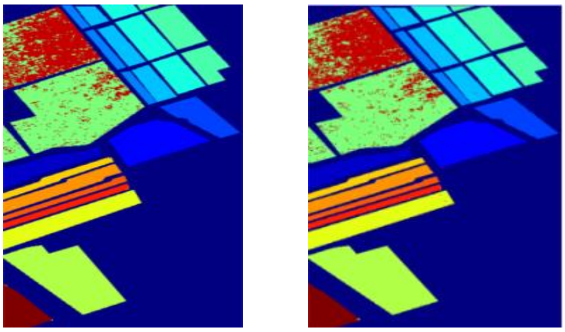
The Ground truth of Indiana Pine

Fig.7. (The reproduced (GT) by the proposed hybridizations for 30 bands of Indiana Pine with its reel GT



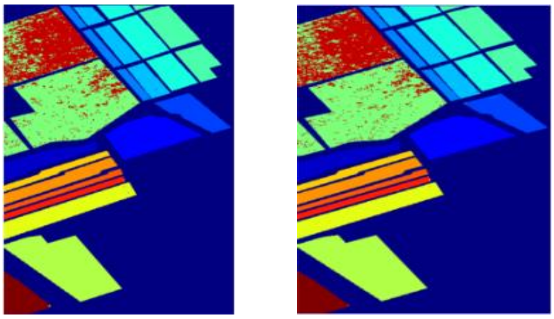
The reproduced GT of Salinas by CMI_GA

The reproduced GT of Salinas by MRLR_GA



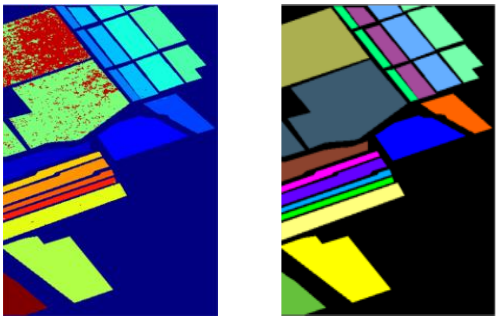
The reproduced GT of Salinas by mRMR_GA

The reproduced GT of Salinas by NMIFS_GA



The reproduced GT of Salinas by MIFS_U2_GA

The reproduced GT of Salinas by MIFSU_GA



The reproduced GT of Salinas by MIFS_GA

The Ground truth of Salinas

Fig.8. (The reproduced (GT) by the proposed hybridizations for 30 bands of SALINAS with its reel GT

V. CONCLUSION

This work combined two parallel works. First, a hybrid method was proposed between the filters based on mutual information MIFs (NMIFS, MIFS, MIFS-U and mRMR, MRLR, CMI) with the genetic algorithm to benefit from these two advantages: speed and efficiency. And as the second aim of this work is to study the feasibility of this hybridization while reporting their behavior in the selection of hyperspectral images.

The first proposal was the CIM and genetic algorithms as a selection strategy. This hybridization has proven with distinction its efficiency in the selection of hyperspectral images, and it comes back to its balance between the relevant and the redundant in its formula. All the results of the selected band classification are calculated by two classifiers RBF-SVM and the KNN. CMI_GA reached over 95% with just 30 bands from Salinas and more than 93% for 30 bands from Indiana and 25 bands from Pavia. The performance of CMI_GA and MRLR_GA overcame the limitations of other parameterized hybridization (PH) based on the parameterization of incremental filters such as NMIFS, MIFS, MIFS-U, MIFS-U2, and even mRMR. However, this latest hybridization failed to define the dependency relationship between the bands due to several constraints that were carefully cited, which explains its failure in these tasks. The analytical and comparative study made deduced that the feasibility of using all of this hybridization is applicable with the privilege of not including filters that have PH. they are less efficient and impractical for handling this type of data. Above all, such a problem does not require human intervention during treatment.

ACKNOWLEDGMENT

Many thanks to M Graña, MA Veganzons, B Ayerdi for hyperspectral Dataset from Grupo De inteligencia Computacial (GIC).

REFERENCES

[1] van Ruitenbeek, F. J., Bakker, W. H., van der Werff, H. M., Zegers, T. E., Oosthoek, J. H., Omer, Z. A., ... & van der Meer, F. D., Mapping the wavelength position of deepest absorption features to explore mineral diversity in hyperspectral images. Planetary and space science, 101, (2014) 108-117.

[2] Karakaya, A., & Yüksel, S. E., Target detection in hyperspectral images. In 2016 24th Signal Processing and Communication Application Conference (SIU) (2016) 1501-1504 . IEEE.

[3] Brook, A., Ben-Dor, E., & Richter, R., Fusion of hyperspectral images and LiDAR data for civil engineering structure monitoring. In 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (2010) 1-5 , IEEE.

[4] Zhang, T., Huang, Y., Reddy, K. N., Yang, P., Zhao, X., & Zhang, J., Using Machine Learning and Hyperspectral Images to Assess Damages to Corn Plant Caused by Glyphosate and to Evaluate Recoverability. Agronomy, 11(3) (2021) 583.

[5] Panasyuk SV, Yang S, Faller DV, Ngo D, Lew RA, Freeman JE, Rogers AE. Medical hyperspectral imaging to facilitate residual tumor identification during surgery. Cancer biology & therapy. 1;6(3) (2007) 439-46.

[6] G.F. Hughes, On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Infirmation Theory, 14(1) (1968) 55-63.

[7] Khodr, J., & Younes, R, Dimensionality reduction on hyperspectral images: A comparative review based on artificial data. In 2011 4th International Congress on Image and Signal Processing , 4 (2011) 1875-1883. IEEE.

[8] Ayesha, S., Hanif, M. K., & Talib, R., Overview and comparative study of dimensionality reduction techniques for high dimensional data. Information Fusion, 59 (2020) 44-58.

[9] Du, Q., & Younan, N. H., Dimensionality reduction and linear discriminant analysis for hyperspectral image classification. In International Conference on Knowledge-Based and Intelligent

Information and Engineering Systems Springer, Berlin, Heidelberg. (2008) 392-399.

[10] Du, Q., Modified Fisher's linear discriminant analysis for hyperspectral imagery. IEEE Geoscience and remote sensing letters, 4(4) (2007) 503-507.

[11] Datta, A., Ghosh, S., & Ghosh, A., PCA, kernel PCA, and dimensionality reduction in hyperspectral images. In Advances in Principal Component Analysis . Springer, Singapore. (2018) 19-46.

[12] Lennon, M., Mercier, G., Mouchot, M. C., & Hubert-Moy, L., Independent component analysis as a tool for dimensionality reduction and the representation of hyperspectral images. In IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217) 6 (2001) 2893-2895. IEEE.

[13] Dobigeon, N., Tourneret, J. Y., Richard, C., Bermudez, J. C. M., McLaughlin, S., & Hero, A. O., Nonlinear unmixing of hyperspectral images: Models and algorithms. IEEE Signal processing magazine, 31(1) (2013) 82-94.

[14] Cao, F., Yang, Z., Ren, J., Jiang, M., & Ling, W. K., Linear vs. nonlinear extreme learning machine for spectral-spatial classification of hyperspectral images. Sensors, 17(11) (2017) 2603.

[15] Kang, X., Li, S., & Benediktsson, J. A., Feature extraction of hyperspectral images with image fusion and recursive filtering. IEEE Transactions on Geoscience and Remote Sensing, 52(6) (2013) 3742-3752.

[16] Fang, L., He, N., Li, S., Plaza, A. J., & Plaza, J., A new spatial-spectral feature extraction method for hyperspectral images using local covariance matrix representation. IEEE Transactions on Geoscience and Remote Sensing, 56(6) (2018) 3534-3546.

[17] Archibald, R., & Fann, G., Feature selection and classification of hyperspectral images with support vector machines. IEEE Geoscience and remote sensing letters, 4(4) (2007) 674-677.

[18] Serpico, S. B., & Bruzzone, L., A new search algorithm for feature selection in hyperspectral remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 39(7) (2001) 1360-1367.

[19] Santos, A. C. S., & Pedrini, H., A combination of k-means clustering and entropy filtering for band selection and classification in hyperspectral images. International Journal of Remote Sensing, 37(13) (2016) 3005-3020.

[20] Cao, X., Ji, B., Ji, Y., Wang, L., & Jiao, L., Hyperspectral image classification based on filtering: a comparative study. Journal of Applied Remote Sensing, 11(3) (2017) 035007.

[21] Yao, X., & Zhao, C., Hyperspectral anomaly detection based on the bilateral filter. Infrared Physics & Technology, 92 (2018) 144-153.

[22] Tu, B., Yang, X., Li, N., Ou, X., & He, W., Hyperspectral image classification via superpixel correlation coefficient representation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(11) (2018) 4113-4127.

[23] Sawant, S. S., & Prabukumar, M.,. A survey of band selection techniques for hyperspectral image classification. Journal of Spectral Imaging, 9 (2020).

[24] Guo, B., Gunn, S. R., Damper, R. I., & Nelson, J. D., Band selection for hyperspectral image classification using mutual information. IEEE Geoscience and Remote Sensing Letters, 3(4) (2006) 522-526.

[25] Datta, A., Ghosh, S., & Ghosh, A., Wrapper-based feature selection in hyperspectral image data using self-adaptive differential evolution. In 2011 International Conference on Image Information Processing (2011) 1-6. IEEE

[26] Merzouqi, M., Sarhrouni, E., & Hammouch, A., Classification and reduction of hyperspectral images based on the motley method. In Proceedings of the Computational Methods in Systems and Software Springer, Cham. (2019) 155-164.

[27] Sarhrouni, E., Hammouch, A., & Aboutajdine, D., Dimensionality reduction and classification feature using mutual information applied to hyperspectral images: a wrapper strategy algorithm based on minimizing the error probability using the inequality of Fano. arXiv preprint arXiv:1211.0055. (2012).

[28] Loggenberg, K., & Poona, N., A Feature Selection Approach for Terrestrial Hyperspectral Image Analysis. South African Journal of Geomatics, 9(2) (2020) 302-320.

[29] Boggavarapu, L. P. K., & Manoharan, P., A new framework for hyperspectral image classification using Gabor embedded patch-based convolutional neural network. Infrared Physics & Technology, 110 (2020) 103455.

[30] Maria, M., El Kebir, S., & Ahmed, H., Reduction dimensionality of hyperspectral imagery using genetic algorithm and mutual information and normalized mutual information as a fitness function. International Review of Applied Sciences and Engineering, 12(1) (2021) 64-75.

[31] Guo, B., Gunn, S., Damper, B., & Nelson, J., Adaptive band selection for hyperspectral image fusion using mutual information. In 2005 7th International Conference on Information Fusion 1 (2005) 8. IEEE.

[32] Verdú, S., Generalizing the Fano inequality. IEEE Transactions on Information Theory, 40(4) (1994) 1247-1251.

[33] Kwak, N., Feature extraction is based on the direct calculation of mutual information. International Journal of Pattern Recognition and Artificial Intelligence, 21(07) (2007) 1213-1231.

[34] Banit'ouagua, I., Kerroum, M. A., Hammouch, A., & Aboutajdine, D., Band selection by mutual information for hyperspectral image classification. International Journal of Advanced Intelligence Paradigms, 8(1) (2016) 98-118.

[35] Battiti, R., Using mutual information for selecting features in supervised neural net learning. IEEE Transactions on neural networks, 5(4) (1994) 537-550.

[36] Kwak, N., & Choi, C. H., Input feature selection by mutual information based on the Parzen window. IEEE transactions on pattern analysis and machine intelligence, 24(12) (2002) 1667-1671.

[37] [37] Huang, J., Cai, Y., & Xu, X., A wrapper for feature selection based on mutual information. In 18th International Conference on Pattern Recognition (ICPR'06) 2 (2006) 618-621 . IEEE.

[38] Estévez, P. A., Tesmer, M., Perez, C. A., & Zurada, J. M., Normalized mutual information feature selection. IEEE Transactions on neural networks, 20(2) (2009) 189-201.

[39] Lee, S., Park, Y. T., & d'Auriol, B. J. (2012). A novel feature selection method based on normalized mutual information. Applied Intelligence, 37(1), 100-120.

[40] Peng, H., Long, F., & Ding, C., Feature selection is based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on pattern analysis and machine intelligence, 27(8) (2005) 1226-1238.

[41] Lee, S. M., Islam, M. M., Kim, J., Kim, Y. H., Jeong, I., & Kim, J. M., Filter and Wrapper-based Feature Selection Using Mutual Information for Rolling Elements Bearing Diagnosis. In Proceedings on the International Conference on Artificial Intelligence (ICAI) . The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldCom). (2017) 187-190.

[42] Cheng, H., Qin, Z., Qian, W., & Liu, W., Conditional mutual information-based feature selection. In 2008 International Symposium on Knowledge Acquisition and Modeling (2008) 103-107 . IEEE.

[43] Holland, J. H., Adaptation in natural and artificial systems: an introductory analysis of biology, control, and artificial intelligence applications. MIT press. (1992).

[44] Sheikh, R. H., Raghuwanshi, M. M., & Jaiswal, A. N., Genetic algorithm-based clustering: a survey. In First International Conference on Emerging Trends in Engineering and Technology (2008) 314-319 . IEEE.

[45] Nhaila, Hasna, et al. Supervised classification methods applied to airborne hyperspectral images: a comparative study using mutual information. Procedia computer science 148 (2019) 97-106.