Clustering-based Unsupervised Hyperspectral Band Selection using Single-Layer Neural Networks

Mateus Habermann, Vincent Frémont, Elcio H. Shiguemori

Clustering-based Unsupervised Hyperspectral Band Selection using Single-Layer Neural Networks

Mateus Habermann[1,2] Vincent Frémont[1] Elcio Hideiti Shiguemori[2]

[1]Sorbonne Universités, Université de Technologie de Compiègne,
CNRS, Heudiasyc UMR 7253, CS 60319, 60203 Compiègne cedex, France.
[2] Instituto de Estudos Avançados (IEAv), Força Aérea Brasileira, Brasil.

mateus.habermann@hds.utc.fr

Résumé

Les images hyperspectrales fournissent des détails de la scène observée en exploitant les bandes spectrales continues. Cependant, le traitement de ces images est long à cause de leurs grandes dimensions. Donc, la sélection des bandes est une pratique commune qui est adopteé avant qu'un traitement soit fait. Ainsi, dans ce travail une nouvelle approche non-supervisé pour la sélection des bandes, basée sur le clustering et les réseaux de neurones, est proposée. Une comparaison avec quatre autres aproches de sélection de bandes montre la validité de l'algorithme proposé.

Mots Clef

Sélection de bandes, non-supervisé, clustering.

Abstract

Hyperspectral images provide fine details of the observed scene by exploiting contiguous spectral bands. However, the processing of such images turns out to be heavy, due to the high dimensionality. Therefore, band selection is a common practice that has been adopted before any processing takes place. Thus, in this work, a new unsupervised approach for band selection based on clustering and neural network is proposed. A comparison with four other band selection approaches shows the validity of the proposed framework.

Keywords

Band selection, unsupervised, clustering.

1 Introduction

Hyperspectral images (HSI) are composed of many continuous bands of the electromagnetic spectrum [1]. Thus, a broad range this spectrum can be covered, which provides lots of information about the scene under analysis. Consequently, those images are useful to Remote Sensing (RS), enabling tasks such as image classification and target detection.

Nevertheless, a large amount of bands can bring problems for data processing, storage and transmission. Moreover, the high dimensionality of HSI can cause the curse of dimensionality [2]. Those facts may cause difficulties in many RS applications, thus, dimensionality reduction is normally adopted.

When it comes to dimensionality reduction, *feature extraction* (FE) is a commonly used method. According to the FE methodology, new features are created by combining the original ones. The resulting features lie in lower dimension, and still they keep much of the original information. One very popular FS technique is Principal Component Analysis (PCA). Since FS changes the original image information, it should not be used in cases that demand the physical meaning of individual bands [3].

Band selection (BS) is another approach for dimensionality reduction. The dimension of the feature space is decreased, whereas the original band information is kept [4].

BS methods are classified in two types : *supervised* and *unsupervised*. Normally, supervised BS approaches yield better results compared to unsupervised ones [5]. But, supervised methods need labeled training samples, which are expensive to be collected. Therefore, unsupervised methods are a wise alternative for band selection.

Since unsupervised BS methods don't have class labels at their disposal, the band selection is made taking into account the data structure [6].

K-Means algorithm is an unsupervised technique that clusters data samples according to the data set structure in the feature space [31].

This paper proposes an easily implementable unsupervised BS framework using K-Means clustering and single-layer neural networks (SLN). Initially, the whole data set is clustered by K-Means into two groups. Then, a SLN is used to find a hyperplane that separates those two clusters, and the bands linked to the biggest and smallest coefficients of the hyperplane

equation are selected. This procedure is considered as one iteration of the method, and two bands are selected at each iteration. If more bands are needed, another iteration is started using either of the clusters just generated, instead of the whole data set.

The contributions of this work are summarized as follows : *i)* It is a novel band selection framework based on K-Means clustering and single-layer neural networks ; *ii)* We compare the proposed method with other unsupervised band selection algorithms.

The rest of the paper is organized as follows : In Section 2, a literature review is presented. Section 3 describes the proposed method. In Section 4, one can find the dataset, classifiers used, the competitors and the results. Finally, Section 5 provides a conclusion to his paper.

2 State-of-the-art approaches

The proposed band selection framework is unsupervised, so only recent state-of-the-art unsupervised approaches are cited in this Section. In the literature, one finds lots of works addressing BS following many different perspectives and methodologies.

Due to the high HSI dimensionality, the different classes present in the image may lie in manifolds embedded in subspaces of the original feature space. Furthermore, it is also possible to explore the sparsity of the data set in order to find a more meaningful data representation. For example, in [14], the authors propose a new method in which they look for salient bands. The number σ of selected bands is user-defined. Then, the band selection algorithm has two steps. Firstly, β bands are selected by means clone selection algorithm, which seeks to minimize the Euclidean distance amongst elements of the same class, whereas maximizing the distance of elements from different classes. After that, if $\beta < \sigma$, those β bands already chosen will serve as seeds to a Manifold Ranking (MR) algorithm. MR sorts the remaining bands, and the most dissimilar band is added to the β group. This step is repeated until $\beta = \sigma$. In [15], the authors propose a BS framework based on sparsity. Initially, the most representative bands are obtained according to the correlation matrix, whereas the block-diagonal structure is measured to segment bands into subspaces. Then, a method for band selection based on trace LASSO and spectral clustering is used. In [16], the authors propose a method that initially represents data instances as sparse coefficient vectors by solving a L2-norm optimization using the least squares regression (LSR) algorithm. Then, a correct segmentation of band vectors is made using the resulting LSR matrix with sparse and block-diagonal structure. After that, a similarity matrix is constructed by angular similarity measurement, and then the size of the band subset is calculated by the distribution compactness plot algorithm. In [17], the authors state that all HSI bands can be represented by a band subset. Thus, they propose a sparse representation of bands with row-sparsity constraint. Besides, a dissimilarity-weighted regularization term is integrated with the self-representation model, to avoid contiguous bands. The problem is solved by the alternating direction method of multipliers, and the representative bands can be chosen. In [18], a fast and robust self-representation framework to select a band subset is proposed. It is assumed the separability structure of the spectral bands, thus the problem may be seen as non-negative matrix factorization. After that, an optimizing convex problem is addressed and augmented Lagrangian multipliers are used to select the band subset. In [19], the authors propose a BS framework that can capture the inter-band redundancy through low-rank modeling. Then, by using an affinity matrix and concepts of data quality, the most representative bands are selected. In [20], a BS method based on column subset selection is proposed. By means of column subset selection problem, it is possible to select some bands maximizing the volume of the selected subset of columns. The high dimensionality decreases the contrast amongst bands, thus Manhattan distance is used to get a higher quality in the BS process. In [21], the authors propose a framework that removes low-discriminating bands that normally need to be discarded manually. Based on the spatial structure of the data set, it is possible to determine which bands have low-discriminating power. Then, a new clustering algorithm is proposed in order to define the optimum number of bands to be selected.

Another criterion that can be used in BS strategies is the HSI data information analysis. For example, in [22], the authors propose a framework that integrates both the overall accuracy and redundancy. Thus, an optimization problem using adaptive balance parameter is devised to handle the trade-off between the overall accuracy and redundancy. Furthermore, an unsupervised overall accuracy prediction method was adopted. In [23], the authors propose a framework that merges the concept of noiseadjusted principal components with maximum determinant of covariance matrix. A new index to measure the HSI quality is also proposed, taking into account signal-to-noise ratios (SNR) and correlation of bands. Based on the new index, the authors devise an unsupervised band selection method, which considers the quality of the data set as selection criterion. It selects bands with both high SNR and low correlation. In [24], the authors propose a BS method based on the dissimilarity amongst neighboring bands. They use an intermediary representation named spectral rhythm, which can take advantage of a pixel sampling strategy, what ends up improving its efficiency without reducing the selected bands quality. In [25], the authors propose a method

based on information-assisted density peak index. It takes into account the intraband information entropy into the local density and intercluster distance to ensure cluster centers with a high quality. Besides, the channel proximity and band distance are integrated to control the local density compactness. The bands with top-ranked scores may get clear global distinction, good local density and also high informative quality. In [26], the authors formulate the BS as a channel capacity problem. After constructing a band channel with the original bands. Then, some bands are selected by Blahut's algorithm, which iteratively finds a feature space that provides the best channel capacity. Thus, neither band prioritization nor interband decorrelation are required. Two iterative methods are devised to find the best band subset, which avoid an exhaustive search.

Using graph theory, in [27] the authors propose a multigraph determinantal point process (MDPP). The aim is to capture the structure amongst bands and find the optimal band subset. For this, multiple graphs are designed to capture the intrinsic relationship amongst bands. Besides, the proposed MDPP is used to model the multiple dependencies in graphs, providing an efficient search strategy for the BS process.

Evolutionary computation with optimization have been largely used by BS methods. For example, in [9], the authors propose an incorporated rank-based multiobjective band selection framework, to avoid conflicting objective functions, such as Jeffreys-Matusita (JF) and Bhattacharyya distances. During the processing, the spectral bands are transformed into binary vectors, whose elements are subjected to flipping with a certain probability. In [10], the authors propose a framework that handles two conflicting objective functions. One function is designed to represent the information contained in the selected bands, by means of entropy. The other function is set as the number of selected bands. Both objective functions are optimized simultaneously by a multiobjective evolutionary algorithm. In [11], the authors propose a framework for band selection which employs two objective functions using JF. During the search process, the spectral bands are treated as firefly variables. In [12], a framework for band selection based on fuzzy clustering and swarm optimization in proposed. The authors devise a modified fuzzy clustering method for band selection, whose drawbacks are alleviated by swarm optimization. In [13], the authors proposed a BS framework based on memetic algorithms (MA). Firstly, MA is used to select a subset of spectral bands. Also, a objective function is designed to select bands considering both bands information and redundancy deduction. The authors claim that this method is not only computationally faster than exhaustive search approaches, but also has comparable performances.

Finally, clustering techniques can also be used in band selection methods. For instance, in [28] the authors propose a framework that removes redundancy amongst bands by means of clustering. Then, from each cluster one representative band is selected. After that, the bands are ranked according to their classification capabilities. In [29], the authors propose a framework based on dual clustering that takes into account the contextual information. For this, a novel descriptor that reveals the image context is devised, in order to select the representatives of each cluster, taking into consideration the mutual effects of each cluster.

This paper uses K-Means to generate clusters according to intrinsic structure of the data set in the feature space, taking into account only the spectral information.

3 Proposed Framework

3.1 Definitions

Let $C^{(0)}$ be the whole data set corresponding to a hyperspectral image, whose elements are vectors $x_i \in \mathbb{R}^{1\times d}$ that contain spectral signatures, where d is the number of bands.

Let S be the set containing the selected bands, and G the set with bands highly correlated to those in S. Let A be the set that has the original spectral bands a_k, with $k = 1, 2, ..., d$. And let σ be a previously determined quantity of bands to be selected.

Let $f : F \longrightarrow L$ be a single-layer neural network, and F is the feature space comprised by $A \setminus (S \cup G)$, and $L = \{0, 1\}$. The input to f is a vector x and its output is a scalar defined by

$$\hat{L} = f(z) = \frac{1}{1 + e^{-z}}, \tag{1}$$

with $z = W^T x + b$, where W and b are the weights and bias of the neural network, respectively.

According to (1), $\hat{L} \in [0, 1]$, and to give it a binary value, these criteria are used :

— if $z < 0 \implies f < 0.5 \implies \hat{L} \leftarrow 0$,
— if $z \geq 0 \implies f \geq 0.5 \implies \hat{L} \leftarrow 1$.

Therefore, the signal of z indicates whether an input sample will be assigned to class 0 or 1.

Since the input data is normalized into $[0, 1]$, the weights $w_l \in W$, with $l = 1, ..., p$, in the equation

$$z = x_i^1 w_1 + x_i^2 w_2 + ... + x_i^p w_p + b \tag{2}$$

cause a strong impact in determining the signal of z, and, consequently, the estimate $\hat{L}_i$ for x_i. In (2), p represents the cardinality of $A \setminus (S \cup G)$.

The cost function of f is cross-entropy. The training is conducted by using stochastic gradient descent and the back-propagation algorithm.

Lastly, let $C_g^{(l)}$ be a partition of $C^{(0)}$, where l is the partition level, and $C_1^{(l)} \cup C_2^{(l)} \cup \cdots \cup C_g^{(l)} = C^{(0)}$, and $C_p^{(l)} \cap C_q^{(l)} = \emptyset, \forall p \neq q$. There may be several levels, that is, $l = 1, 2, 3, \ldots$, and for each level the number of partitions g is given by $g = 2^l$.

3.2 Description

General view. The proposed framework begins with an empty subset of selected bands, that is, $S = \emptyset$, to which the bands selected from A will be added. At the first iteration, $C^{(0)}$ is split by the K-Means algorithm into two partitions, $C_1^{(1)}$ and $C_2^{(1)}$. Then, we use a single-layer neural network to find a hyperplane that separates those two partitions. After that, two bands are selected, and consequently discarded from the data set. If more bands are needed, we keep repeating this procedure in deeper levels.

As the proposed method is based on both clustering and single-layer neural networks, we shall call it `CSLN`. Its characteristics are described below.

Iterations. `CSLN` is an iterative band selection method. At each iteration, a binary classification problem between K-Means-generated partitions $C_p^{(l)}$ and $C_q^{(l)}$ is to be solved by the function f. Since two bands are selected at each iteration, one needs to repeat the process until the desired number of bands σ is attained. When σ is even, one needs $\sigma/2$ iterations. If σ is an odd number, $(\sigma + 1)/2$ iterations are necessary, and the first σ selected bands are kept.

Selection of bands. After the training of the neural network, it is possible to give degrees of importance to all $a_k \in A \setminus (S \cup G)$. As every element $x^l \in x$ is directly linked to w_l, for $l = 1, ..., p$, the magnitude of w_l is a indicator for the band a_l. In (2), the largest and the smallest weights constitute the most important contributions to the signal of z. Thus, the bands linked to those weights are also considered the most important, and, consequently they are added to the set S. The feature space F can be updated by $A \setminus (S \cup G)$.

Avoiding highly correlated bands. The bands of a hyperspectral image are contiguous, which causes a high correlation among neighboring bands.

Bearing this in mind, we adopt a method that avoids the selection of highly correlated bands. For each band $a_k \in F$ we construct a vector v_k, whose elements are the bands indices in a descending order in relation to the correlation to the band a_k. Thus, the following procedure is adopted :

- At a certain iteration, a band a_k is selected, so $S \leftarrow a_k$;
- $G \leftarrow a_{v_k(1)}$, and G is, at the beginning, an empty set ;
- After this iteration, the feature space is updated by $A \setminus (S \cup G)$.

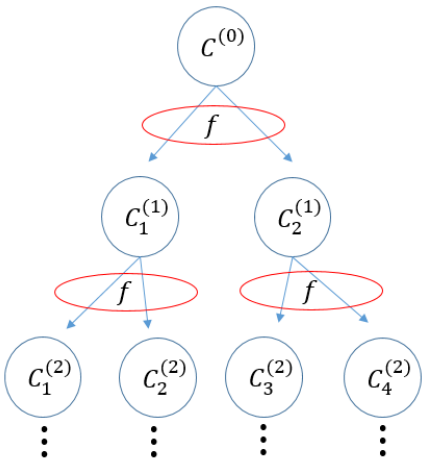


FIGURE 1 – Flowchart of the proposed BS framework. At each binary clustering, a single-layer neural net f is used to select the bands.

We emphasize that only $a_k \in S$ are the selected bands. The bands $a_{v_k(1)} \in G$ are discarded.

Algorithm 1 gives the steps followed by the proposed $CSLN$ framework.

Algorithm 1 Proposed band selection framework

1: Input : $C^{(0)}$, A, σ, $S = \emptyset$ and $G = \emptyset$
2: **for** $r = 1$: maxIterations **do**
3: Use the function f to find a hyperplane that separates $C_p^{(l)}$ and $C_q^{(l)}$ clustered by K-Means
4: Select the bands $a_k \in A$ related to the largest and smallest $w \in W$
5: $S \leftarrow a_k$, and $G \leftarrow a_{v_k(1)}$
6: Update the feature space F by $A \setminus (S \cup G)$
7: Return : S

Fig. 1 depicts the proposed framework. Initially, the whole data set is split into two clusters by K-Means. Then, a single-layer neural network f is used to find a hyperplane that separates the clusters. This process is repeated until the desired number of bands is selected.

4 Results

The results of the proposed method are exhibited in this Section. Moreover, they are compared with other band selection approaches by considering the accuracy of two supervised classifiers, namely, Classification and Regression Trees (CART) and K-Nearest Neighbors (KNN). The input data are the selected bands.

The image used is the Indian Pines. It consists of 145×145 pixels and 224 spectral reflectance bands in the $0.4 - 2.5$ µm wavelength range. As for the ground

truth, we count on 16 classes, and they are used only for classification comparison purposes.

4.1 Competitors

The performance of the proposed method is compared with four other BS approaches.
The first method is also clustering-based [32], and it will be referred to as `WaLuDi`. The second approach uses both ranking and clustering for band selection [33], and we will call it `CR`. The third competitor resorts to band elimination with partitioned image correlation [34], and it will be referred to as `EM`. Lastly, the fourth competitor relies on information divergence, and this method will be called `ID` [35].
As already stated in Section 3.2, our proposed method will be referred to as `CSLN`.

4.2 Selected bands

The bands selected by the proposed framework are displayed in Table 1. We have only the first 18 best-ranked bands of our competitors, thus the analyses of results are restricted to this number of bands.
The bands in Table 1 are sorted according to the order they were selected. For example, at the first iteration, the bands 2 and 42 were selected.
The results comparisons are made with different quantities σ of selected bands, that is, $\sigma_s = s \times 3$, with $s = 1, 2, 3, 4, 5, 6$. Thus, for $\sigma_2 = 6$, for example, the first six bands of Table 1 are used.

TABLE 1 – The selected bands according to the order of selection by the proposed method.

Selected bands	2, 42, 6, 39, 22, 58, 25, 62, 71, 101, 94, 151, 111, 203, 156, 183, 171, 215.

4.3 Results comparison

The classification results exhibited throughout this paper are the mean values over ten runs. The standard-deviation values are also displayed.
In Table 2, the results of the KNN classifier are shown. The proposed method `CSLN` has the best results using 3, 6 and 9 bands. It is illustrated in Fig. 2 (a).
In Table 3, the overall results achieved by the CART classifier are exhibited. The proposed framework achieves the best results with 3 and 9 bands. Fig. 2 (b) provides a visual perspective of the results.

4.4 Remarks about the results

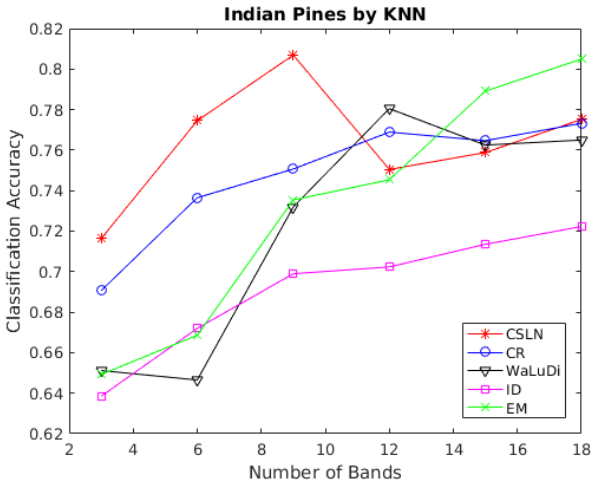
KNN results are, in general, superior than that of CART : 73.26% and 63.13%, respectively. This may be attributed to the fact that CART splits the feature space into regions that correspond to the classes. Therefore, if x_i is found in a region corresponding to a class α, for example, it will be classified as α, even if it belongs to class β. Whereas, in this same situation, KNN would analyze the K nearest neighbors of x_i before assigning it a label. Consequently, KNN outperforms CART when the class boundaries are highly non-linear. Moreover, according to Tables 2 and 3, one can notice that, in general, the accuracies get better as more bands are used. Both facts are depicted in Fig. 3.
As for the band selection methods, using the KNN classifier, the proposed BS framework achieves the best results in three situations. Furthermore, by analyzing the standard-deviation values of Table 2, one can notice that the `CSLN` method have statistically better results. With regard to the CART classifier, our method has the best accuracy with 3 and 9 bands. With 6, 12, 15 and 18 bands, `CSLN` has similar performance in relation to its best competitors.

5 Conclusion

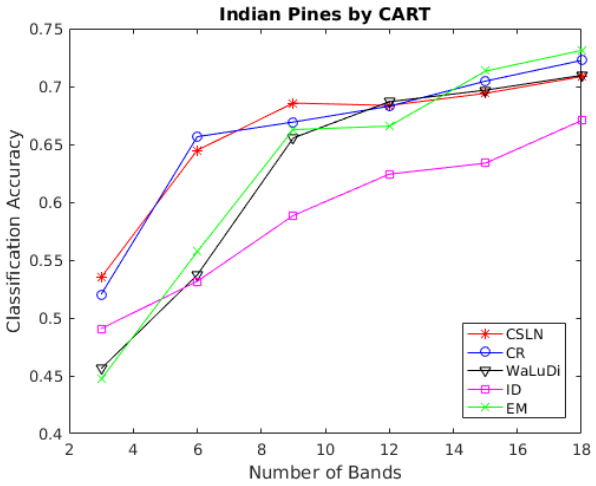
The fine spectral details provided by HSI allow for a good characterization of objects in the scene. But, the large amount of spectral information can also bring inconveniences in terms of processing and storage. Therefore, this paper proposed a band selection framework to decrease the image dimensionality.
The proposed unsupervised BS method is based on K-Means clustering and single-layer neural networks. It started by clustering the whole data set into two groups. Then, a single-layer neural network was used to find a separating hyperplane between the clusters. The bands linked to the biggest and smallest coefficients of the hyperplane equation were selected. Then, we kept repeating this procedure using the generated clusters to select the desired number of bands.
By analyzing the results, one can see that the KNN classifier outperforms CART for the Indian Pines image. Moreover, the more bands, the better the classifier accuracy. We emphasize that 3 up to 18 selected spectral bands were investigated. When it comes to the proposed method, the best results were achieved in three situations out of six using KNN. With CART, the proposed method got the best results in two situations out of six.
With regard to the future works, we will investigate other clustering algorithms and binary classifiers and use them in our framework.

Acknowledgements

This work was carried out in the framework of the Labex MS2T and DIVINA challenge team, which were funded by the French Government, through the program Investments for the Future managed by the National Agency for Research (Reference ANR-11-IDEX-0004-02).
We are also thankful for the support provided by Brazilian Air Force and Institute for Advanced Studies.



(a)



(b)

FIGURE 2 – Indian Pines image classification results by KNN classifier. In (a), KNN results. In (b), results achieved by CART.

TABLE 2 – KNN results.

	3 bands		6 bands		9 bands		12 bands		15 bands		18 bands	
Method	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
CSLN	**71.63%**	0.66%	**77.45%**	0.63%	**80.69%**	0.38%	75.05%	0.66%	75.88%	0.65%	77.52%	0.37%
WaLuDi	65.12%	1.02%	64.65%	0.25%	73.19%	0.72%	**78.05%**	0.56%	76.25%	0.19%	76.50%	0.67%
CR	69.06%	0.52%	73.65%	1.03%	75.07%	1.43%	76.89%	1.07%	76.47%	1.14%	77.32%	0.22%
EM	64.92%	1.15%	66.86%	1.03%	73.54%	0.28%	74.54%	1.07%	**78.92%**	0.41%	**80.50%**	0.52%
ID	63.85%	079%	67.20%	0.22%	69.90%	0.18%	70.23%	1.16%	71.35%	0.47%	72.23%	1.34%

TABLE 3 – CART results.

	3 bands		6 bands		9 bands		12 bands		15 bands		18 bands	
Method	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
CSLN	**53.51%**	0.90%	64.49%	0.43%	**68.56%**	0.59%	68.36%	1.08%	69.41%	0.48%	70.85%	1.49%
WaLuDi	45.62%	1.00%	53.71%	1.23%	65.55%	0.95%	**68.68%**	0.28%	69.68%	0.75%	70.96%	1.15%
CR	52.03%	1.14%	**65.66%**	0.39%	66.93%	0.37%	68.29%	1.48%	70.46%	0.99%	72.25%	1.91%
EM	44.72%	0.93%	55.72%	1.04%	66.28%	0.52%	66.57%	1.24%	**71.33%**	0.76%	**73.12%**	0.51%
ID	49.07%	0.82%	53.16%	1.35%	58.85%	1.42%	62.43%	1.67%	63.37%	1.01%	67.06%	0.87%

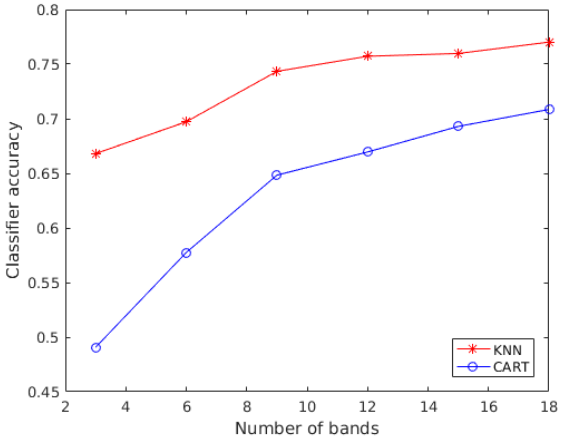


FIGURE 3 – Mean results of all BS methods together.

Références

[1] C. I. Chang, *Hyperspectral Imaging : Techniques for Spectral Detection and Classification*. Springer, 2003, vol. 1.

[2] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63, January 1968.

[3] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in *2014 Science and Information Conference*, Aug 2014, pp. 372–378.

[4] J. Li and H. Liu, "Challenges of feature selection for big data analytics," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 9–15, Mar 2017.

[5] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 1, pp. 138–142, Jan 2011.

[6] B. Chandra and R. K. Sharma, "Exploring autoencoders for unsupervised feature selection," in *2015 International Joint Conference on Neural Networks (IJCNN)*, July 2015, pp. 1–6.

[7] S. S. Haykin, *Neural networks and learning machines*, 3rd ed. Upper Saddle River, NJ : Pearson Education, 2009.

[8] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders : Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, Dec. 2010. [Online]. Available : http ://dl.acm.org/citation.cfm ?id=1756006.1953039

[9] X. Xu, Z. Shi, and B. Pan, "A new unsupervised hyperspectral band selection method based on multiobjective optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2112–2116, Nov 2017.

[10] M. Gong, M. Zhang, and Y. Yuan, "Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 544–557, Jan 2016.

[11] H. Su, Q. Li, and P. Du, "Hyperspectral band selection using firefly algorithm," in *2014 6th Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS)*, June 2014, pp. 1–4.

[12] M. Zhang, J. Ma, and M. Gong, "Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 773–777, May 2017.

[13] M. Zhang, J. Ma, M. Gong, H. Li, and J. Liu, "Memetic algorithm based feature selection for hyperspectral images classification," in *2017 IEEE Congress on Evolutionary Computation (CEC)*, June 2017, pp. 495–502.

[14] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1279–1289, June 2016.

[15] J. Wang, K. Zhang, P. Wang, K. Madani, and C. Sabourin, "Unsupervised band selection using block-diagonal sparsity for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2062–2066, Nov 2017.

[16] W. Sun, L. Zhang, B. Du, W. Li, and Y. M. Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2784–2797, June 2015.

[17] L. Gan, J. Xia, P. Du, and Z. Xu, "Dissimilarity-weighted sparse representation for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 1968–1972, Nov 2017.

[18] W. Sun, L. Tian, Y. Xu, D. Zhang, and Q. Du, "Fast and robust self-representation method for hyperspectral band selection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 11, pp. 5087–5098, Nov 2017.

[19] G. Zhu, Y. Huang, S. Li, J. Tang, and D. Liang, "Hyperspectral band selection via rank minimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2320–2324, Dec 2017.

[20] C. Wang, M. Gong, M. Zhang, and Y. Chan, "Unsupervised hyperspectral image band selection via column subset selection," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 7, pp. 1411–1415, July 2015.

[21] X. Cao, B. Wu, D. Tao, and L. Jiao, "Automatic band selection using spatial-structure information and classifier-based clustering," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pp. 4352–4360, Sept 2016.

[22] C. Sui, Y. Tian, Y. Xu, and Y. Xie, "Unsupervised band selection by integrating the overall accuracy and redundancy," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 1, pp. 185–189, Jan 2015.

[23] K. Sun, X. Geng, L. Ji, and Y. Lu, "A new band selection method for hyperspectral image based on data quality," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2697–2703, June 2014.

[24] L. C. B. dos Santos, S. J. F. GuimarÃčes, and J. A. dos Santos, "Efficient unsupervised band selection through spectral rhythms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1016–1025, Sept 2015.

[25] X. Luo, R. Xue, and J. Yin, "Information-assisted density peak index for hyperspectral band selection," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1870–1874, Oct 2017.

[26] C. I. Chang, L. C. Lee, B. Xue, M. Song, and J. Chen, "Channel capacity approach to hyperspectral band subset selection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 10, pp. 4630–4644, Oct 2017.

[27] Y. Yuan, X. Zheng, and X. Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 51–64, Jan 2017.

[28] A. Datta, S. Ghosh, and A. Ghosh, "Clustering based band selection for hyperspectral images," in *2012 International Conference on Communications, Devices and Intelligent Systems (CODIS)*, Dec 2012, pp. 101–104.

[29] Y. Yuan, J. Lin, and Q. Wang, "Dual-clustering-based hyperspectral band selection by contextual analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1431–1445, March 2016.

[30] L. Wang, C. I. Chang, L. C. Lee, Y. Wang, B. Xue, M. Song, C. Yu, and S. Li, "Band subset selection for anomaly detection in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 4887–4898, Sept 2017.

[31] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*, 4th ed. Academic Press, 2008.

[32] A. MartÃĐnez-UsÃŞMartinez-Uso, F. Pla, J. M. Sotoca, and P. GarcÃĐa-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4158–4171, Dec 2007.

[33] A. Datta, S. Ghosh, and A. Ghosh, "Combination of clustering and ranking techniques for unsupervised band selection of hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2814–2823, June 2015.

[34] ——, "Band elimination of hyperspectral imagery using partitioned band image correlation and capacitory discrimination," *International Journal of Remote Sensing*, vol. 35, no. 2, pp. 554–577, 2014. [Online]. Available : https ://doi.org/10.1080/01431161.2013.871392

[35] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1575–1585, June 2006.