Learning to Combine Spectral Indices with Genetic Programming

Juan Felipe Hernández Albarracín
Institute of Computing
University of Campinas
Campinas, SP 13083–852
Email: juan.albarracin@students.ic.unicamp.br

Jefersson Alex dos Santos
Department of Computer Science
Federal University of Minas Gerais
Belo Horizonte, MG
Email: jefersson@dcc.ufmg.br

Ricardo da S. Torres
Institute of Computing
University of Campinas
Campinas, SP 13083–852
Email: rtorres@ic.unicamp.br

***Abstract*—This paper introduces a Genetic Programming-based method for band selection and combination, aiming to support remote sensing image classification tasks. Relying on ground-truth data, our method selects spectral bands and finds the arithmetic combination of those bands (i.e., spectral index) that best separates examples of different classes. Experimental results demonstrate that the proposed method is very effective in pixel-wise binary classification problems.**

I. Introduction

Remote sensing is a field that has been gaining a huge importance in nature sciences research. With the advance of technologies regarding image acquisition in various spectral bands and their public availability, there is a lot that can be done in order to deduce knowledge from the provided data. One common application relies on the use of spectral indexes, which are fundamental in high-level decision making processes as it is possible to associate the spectral bands of a scene with a measure of a chemical or physical property in the environment.

A common approach used to construct effective hyper- and multi-spectral remote sensing image classification systems relies on the selection of appropriate bands and their combination with the objective of providing a better discrimination among the pixel values observed for different classes. Several successful approaches have been proposed aiming at support both band selection and/or combination [1]–[14]. Most of them, however, do not address both problems at the same time.

Band selection-based methods typically use a metric to select the most informative bands to provide a subset that could be the most appropriate. This strategy is based on the fact that some bands may contain noisy or redundant information, being bad candidates to be selected. It is very useful in many applications because it can not only reduce dimensionality, but also preserves relevant original information of the spectral bands. If prior knowledge is available, supervised methods [1]–[4], [15] can be used to achieve good results. Concerning the unsupervised band selection methods, some approaches exploit clustering techniques, in which the first step is to compute a distance measure for each pair of bands. With these metrics, the bands are grouped in disjoint clusters such that bands in a given cluster tend to be similar to each other according to these metrics, and bands in different clusters tend to be dissimilar. After the grouping step, a representative band from each cluster is chosen [5]–[8], [14]. Search-based methods, on the other hand, aim at finding a good set of bands by evaluating subset features. Using exhaustive search strategies to find the best subset is normally unmanageable for this kind of data, however, several sub-optimal search strategies like sequential backward selection [16] or evolutionary techniques are used in this kind of problem [17], [18]. Rank-based methods use metrics, such as entropy, mutual information, and correlation, to sort the bands by their importance in the selection. Some examples of such metrics can be found in [9]–[13].

In this paper, we introduce a novel approach that uses Genetic Programming (GP) for automatic spectral index learning, addressing the pixel-wise classification problem in remote sensing images. The objective is to provide as output a scalar from which it can be deduced to which class a pixel belongs. The proposed method relies on the Genetic Programming evolutionary algorithm to select and combine image bands at the same time. Band combination functions are modeled as individuals of a population and evolves over generations. The best individuals, those related to combination functions that maximize the separation among values of pixels of different classes, are later used to assign classes to testing samples.

Initial experiments on this mater, presented in this paper, comprise binary classification problems considering the band combination function discovered by the GP apparatus. Performed experiments on two widely used datasets demonstrate that the proposed method is effective in classifying pixels given the band combination functions.

II. Background

A. Remote sensing and spectral indices

Remote sensing consists in deducing knowledge regarding the terrestrial surface, taking advantage of images remotely collected in one or more channels of the electromagnetic spectrum. Images of the same scene in different regions of the spectrum (bands) are known as multi- or hyper-spectral images, depending on the quantity and continuity of the bands that comprise them.

Multi-band images provide valuable information for analysis purposes since different materials and chemicals can be identified by means of their *spectral signature*, i.e. the

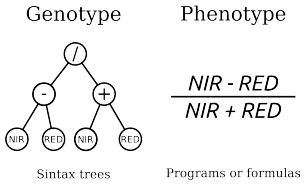


Fig. 1. Computational representation (genotype) of the formula (phenotype) for the NDVI vegetation index.

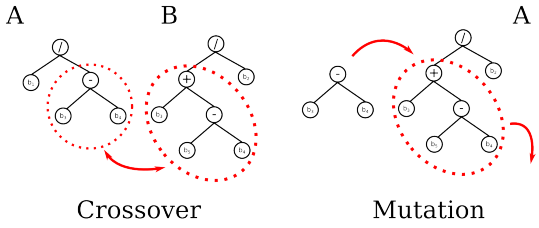


Fig. 2. Crossover and mutation operators for syntax trees.

particular reflectance/emittance of a substance in different wavelengths. *Spectral indices* are functions that associate spectral bands with physical and chemical features of objects of the environment. A relevant example is NDVI (acronym for Normalized Difference Vegetation Index), a vegetation index for biomass measurement, based on the difference between the reflectance of an object in the visible (red) channel and its reflectance in the near-infrared channel:

$$NDVI = \frac{NIR - VIS}{NIR + VIS} \tag{1}$$

where NIR is the emission rate in the near-infrared channel and VIS is the emission rate in the visible (red) channel. Healthy vegetation in most cases presents $NDVI$ values close to 1 or -1, since it reflects big quantities of infrared radiance while absorbing big quantities in visible radiance. Although spectral indices are normally formulated by specialists, various approaches aim to deduce them automatically, relying on ground-truth data and pattern recognition algorithms.

B. Genetic Programming

Genetic Programming (GP) is a machine learning technique that belongs to the family of evolutionary algorithms [19], in which candidate solutions to a complex problem are represented as individuals within a population. A fitness function is defined as a criterion to select and keep those solutions that best solve the problem, according to the Darwinian principle of survival of the fittest.

Particularly in GP, the individuals are computer programs or formulas, whose fitness depends on the result of their execution. Individuals are normally represented as syntax trees and two genetic operations are performed with those trees in order to create new individuals: crossover and mutation. Crossover is the operation with which two trees reproduce, sharing genetic information by exchanging one of their sub-trees. New formed individuals have a chance to mutate, by arbitrarily changing one of their sub-trees by a randomly generated new one. The processes of exploration and exploitation in stochastic search are performed, respectively, with the operations of mutation and crossover. The computational representation of formula is known as the genotype and the formula itself is known as the phenotype of the individual. Fig 1 shows an example of an individual representation associated with the computation of the NDVI vegetation index (see Eq. 1). Figure 2, in turn, illustrates the genetic operators used in the evolutionary process.

The GP algorithm starts with a randomly generated population. Each iteration of the algorithm is considered a generation, in which individuals reproduce, creating new similar ones. The fitness function is used to guide the selection of individuals to reproduce. Different selection approaches may be used [19]. The algorithm keeps iterating until a stop condition is reached, usually defined in terms of the number of generations considered in the evolution process or the level of fitness reached by population individuals.

III. Related work

Various researches have been focusing on the use of GP for remote sensing image band combination. Usually, they aim to find formulas that measure the concentration of chemicals or the presence of an object.

Particularly, the method proposed by Fonlupt et al. in [20] uses GP to measure the concentration of phytoplankton, sediment, and yellow substance in oceans and coastal waters. Using a dynamic fitness function to optimizing for each class at a time, it was possible to outperform traditional polynomial approximations.

Chion et al. proposed in [21] the genetic programming-spectral vegetation index (GP-SVI), a method that evolves a regression model to describe the nitrogen level in vegetation. The fitness depends on the correlation with ground-truth data and the size of the formula. GP-SVI outperformed a considerable amount of regression methods, such as genetic algorithms for partial least squares regression (GA-PLS), multiple regression (MR), tree-based models (TBM), and some classical spectral vegetation indices such as NDVI.

A very similar work was presented few years later by Puente et al. in [22]. That work introduced a genetic programming vegetation index (GPSVI) to estimate the vegetation cover factor in soils to assess erosion. The proposed fitness function depends on the covariance with the cover-management factor obtained by in situ observations.

A slightly different approach is presented by Ross et al. in [23] for mineral classification (3 classes). Binary classifiers are trained with GP and the fitness function depends on the rate of correctly classified examples.

Rauss et al. evolves an index in [24] that returns values greater than 0 when there is grass in the image, and values smaller than 0 in the other case. The fitness function depends on the number of correctly classified examples, relying on ground-truth data.

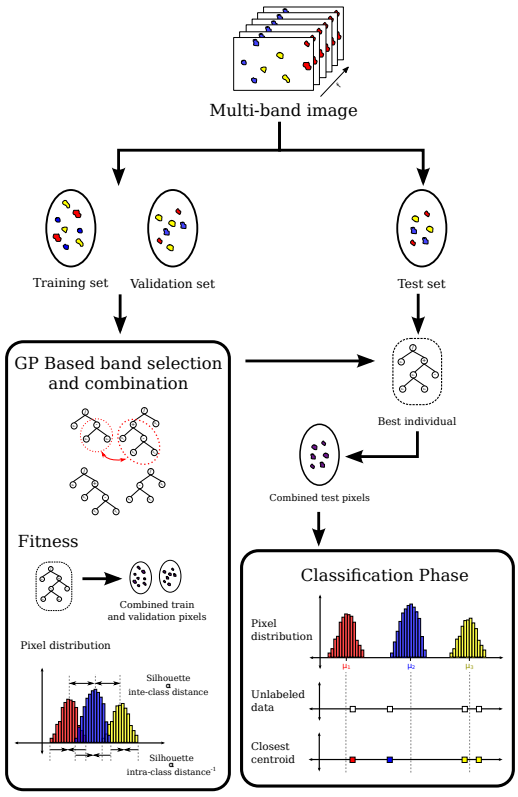


Fig. 3. General pipeline of the proposed method. The labeled pixels of a multi-band image are separated into training, validation, and test sets. Training and validation sets are used in the GP framework to find the individual that best separates the class distributions. Once the GP phase is over, the best individual found is used to combine the bands of the test set and their corresponding pixels are classified according to its distance to the centroids of the classes.

Perkins et al. introduce in [25] GENIE, a hybrid algorithm based on linear genetic programming that finds a combination of successive image processing operations that can reproduce suitable feature planes for conventional classifiers.

Different from the above initiatives, in this paper, we use GP for classification purposes, relying on the quality of the feature space constructed by the band combination that was provided by the algorithm.

IV. Proposed Approach

We propose a technique based on Genetic Programming (GP) for spectral imagery pixel-wise classification, aiming to find a spectral index (i.e., an arithmetic combination of bands) that maximizes the accuracy.

Fig.3 depicts the general pipeline of the proposed method. Given (multi or hyper) spectral imagery with ground-truth data (labeled pixels), our method runs a GP algorithm that evolves formulas used to combine arithmetically the bands of the scenes. Each formula creates a unique gray-scale image, representing the combination of the bands for each pixel. The criteria to evaluate each formula are the inter and intra-class distance of the resulting pixels: those of the same class should be together and far away from pixels of a different class.

Once the best formula is found, its resulting image is used as input to a very simple classifier that, based on the values of the pixels and the available labels, should indicate to what class an unlabeled pixel belongs.

A. The GP-based band selection and combination approach

Each individual (candidate formula) is represented as a syntax tree, whose intern nodes stand for arithmetic operators and the leaves represent both variables b_i (corresponding to each spectral band) and some constants with random values. The fitness function (described below), determines how separable into different classes are the examples in the space defined by one formula. The fitness allows to rank individuals (band combination functions) in order to give the better ones a major chance to be selected for reproduction.

1) Fitness function: Considering each class as a distribution of values of pixels, the measure used to evaluate their separation in a space created by one spectral index is the Silhouette Score introduced by Rousseuw in [26]. The silhouette, used principally for clustering, indicates how well-assigned an object is to its cluster, with respect to the others; it depends on how compact their distributions are and how far from the others.

Let the *average dissimilarity* of an object x to a cluster C_i be the mean of the distances of x to all the objects that belong to C_i. Let $a(x)$ be the average dissimilarity of x to its own cluster and $b(x)$ the smallest average dissimilarity of x to the other clusters, so $b(x)$ indicates how dissimilar x is with respect to the most similar cluster that is not its own. The silhouette of a single object x is defined as.

$$s(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}} \tag{2}$$

yielding values between -1.0 and 1.0. Negative values mean that there is at least one cluster that is more similar to x than the cluster to which it is assigned currently. The larger the silhouette, the better is the assignment of x to its cluster. The overall score of the clustering method is the mean of the silhouettes of all the objects. The objective of the GP framework is to maximize this score.

Although its common use is for parameter tuning of clustering methods, the silhouette is an excellent measure to evaluate feature spaces as well, since, relying on ground-truth data, it indicates how separable are the elements of two or more classes. Fig.4 shows the distributions of a set of pixel values, having the previous knowledge of what class they belong (either the red or blue class). Here we compare the definition of the classes for three different moments in the GP evolution: generation 1 (Fig. 4(a)), 100 (Fig. 4(b)), and 200 (Fig. 4(c)). Note that the separation is more evident for the distributions with the highest silhouette score.

2) Handling overfitting Using Validation Sets: To avoid overfitting, a further selection of individuals was performed, using a validation set, as done by Torres et al. in [27]. The validation consists in keeping the best k individuals through all generations and using each one of them to combine the pixels

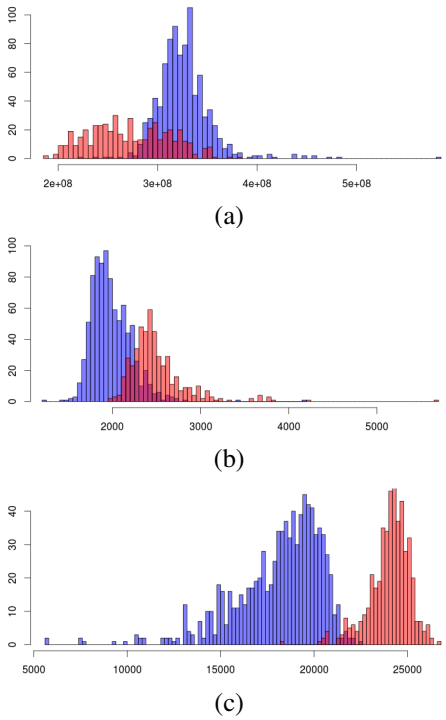


Fig. 4. Comparison of class distributions with different silhouette scores, corresponding to different generations in the evolution. (a) shows the distribution at the end of the first generation, with a silhouette score of 0.64. (b) shows the distribution at the end of generation 100, with a score of 0.83. (c) shows the distribution at the end of generation 200, with a score of 0.84.

of the validation set. Let μ^i be the mean of the silhouette scores for the training and validation sets of individual i, and σ^i its standard deviation. The final score of the individual, S^i, is given by:

$$S^i = \mu^i - \sigma^i \tag{3}$$

The larger the difference between the silhouette for the training and the validation set, the larger σ^i and probably, the smaller the final score. This score will impact negatively those individuals whose performance in the validation set is much smaller than the training set.

B. Classification phase

Counting on a good definition of the classes, provided by the best individual found in the GP phase, the classification task can be done with a very simple algorithm. In this case, we chose the minimum distance to the value of a central tendency measure (centroid) for each sample of training points corresponding to a class. Once the centroids of each class are calculated, an unlabeled sample can be classified into the class of the closest centroid. The central tendency measure used is the mean, since it is the most widely used one. The investigation of other strategies for class assignment based on the trained distributions is left as future work.

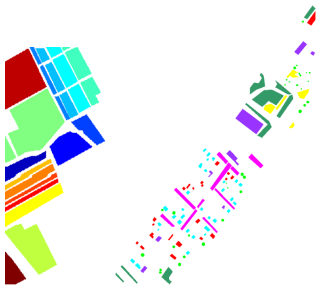


Fig. 5. Ground truth of the datasets used: Salinas (left) and Thetford Mines (right).

TABLE I
CLASSES IN THE SALINAS DATASET AND THE NUMBER OF EXAMPLES PER CLASS

Class	Num.	Class	Num.
Brocoli_green_weeds_1	2009	Brocoli_green_weeds_2	3726
Fallow	1976	Fallow_rough_plow	1394
Fallow_smooth	2678	Stubble	3959
Celery	3579	Grapes_untrained	11271
Soil_vinyard_develop	6203	Corn_senesced_green_weeds	3278
Lettuce_romaine_4wk	1068	Lettuce_romaine_5wk	1927
Lettuce_romaine_6wk	916	Lettuce_romaine_7wk	1070
Vinyard_untrained	7268	Vinyard_vertical_trellis	1807

V. Experimental setup

In this paper, we report results for binary classification problems. The objective is to show that it is possible to find a combination of bands that successfully separates samples of different classes. For each pair of classes, a spectral index is evolved and used to classify testing samples.

A. Datasets

We tested our method on two datasets. The first one is a scene collected over Salinas Valley, California (we will call this dataset just as Salinas) and the second one is a scene collected over an urban area near Thetford Mines in Québec, Canada (we will call this data set just as Thetford Mines). Fig. 5 shows the ground of both datasets. Details of each one will be presented below.

1) Salinas: This scene was collected by the AVIRIS sensor in 224 spectral bands, with a spatial resolution of 3.7 meters/pixel, forming a 512 x 217 image.[1] Its ground truth contains 16 classes, as shown in Table I.

2) Thetford Mines: This scene was collected by a 84-channel sensor, with a spacial resolution of about 1 meter/pixel. The wavelengths covered are between 7.8 and 11.5 micrometers. This is one of the data sets of the 2014 IEEE GRSS Data Fusion Contest.[2] Its ground-truth contains 7

[1]Data retrieved from `http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes`. Last accessed on May 11th, 2016.

[2]`http://www.grss-ieee.org/community/technical-committees/data-fusion/2014-ieee-grss-data-fusion-contest/`. Last accessed on May 11th, 2016.

TABLE II
CLASSES IN THE THETFORD MINES DATASET AND THE NUMBER OF EXAMPLES PER CLASS

Class	Num.	Class	Num.	Class	Num.
road	4293	trees	1027	red roof	1739
grey roof	1973	concrete roof	3797	vegetation	7167
bare soil	1711				

TABLE III
GP PARAMETERS SETUP

Parameter	Value
Population size	100
Generations	200
Operators (intern nodes)	{+, -, *, %}
Parameters (laves)	$\{b_i : 0 \leq i < n\} \cup \{c_j \in [0, 10^6]\}$
Maximum initial tree depth	6
Maximum crossover depth	15
Selection method	Tournament $\times$ 3
Crossover rate	0.9
Mutation rate	0.1
Individuals kept for val.	10

classes, as shown in Table II.

B. Parameters

Table III shows a summary of the configuration of the GP algorithm. The parameters of the formulas (i.e., leaves of the trees) are random real numbers between 0 and 10^6, and the variables that indicate the spectral bands. The possible operators of the formula (i.e., internal nodes) are addition $(+)$, subtraction $(-)$, multiplication (*), and protected division $(\%)$.[3]

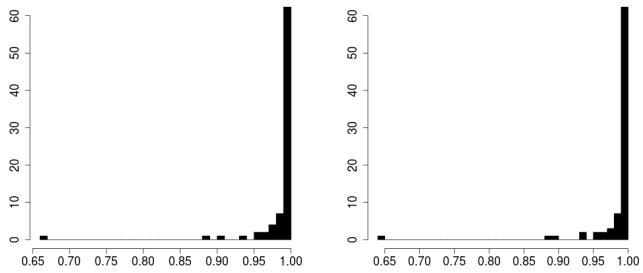
Experiments were performed with 100 individuals and 200 generations. New randomly generated trees will not have a depth greater than 6, and crossover and mutation operations will not yield trees with a depth greater than 15.

The selection method is *tournament with three individuals*. It consists in selecting at random three individuals from the population and allowing the best one to go to crossover, as many times as a new generation is completed. Once two individuals are chosen, they have a probability of 0.9 to cross their genetic information and create new individuals. Every new individual in the population has a probability of 0.1 to mutate.

For validation, the best ten individuals of all the generations are re-ranked according to the score described in (3). The best individuals found with and without validation will be kept for classification, and their performances are compared.

No stop condition different to the maximum number of iterations was considered, since it was not clear in the beginning what was the minimum silhouette score needed to yield a good classification accuracy. We use 200 generations based on empirical experiments.

[3]This operator replaces normal division, as suggested by Koza in [19] to avoid division by zero, by returning 1 as the result of the division, every time a 0 is found in the denominator.



(a) (b)

Fig. 6. Histograms of the classification mean accuracies for the 120 pairs of Salinas' classes. Charts (a) and (b) show the distribution for indices learned without and with validation set, respectively.

TABLE IV
OVERALL MEAN ACCURACIES FOR THE SALINAS DATASET

	Without validation	With validation
Mean accuracy	0.9902559	0.9898045

C. Evaluation protocol

The well-known 5-fold cross validation approach was used to evaluate the performance of the classification. Three of the folds were used for training, one for validation, and one for testing. Five runs were executed in such a way that folds were shifted to be used for training, validation, and testing. In the next section, we report the results concerning the use of the best found band combination functions in binary classification tasks.

VI. RESULTS

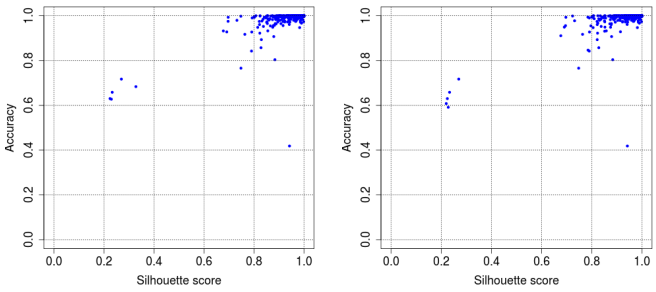
The 5-fold cross validation was performed for each pair of classes in the data sets. In this section, we report the mean of the five accuracies obtained in each experiment. As described above, the experiments were performed considering and not considering the use of the validation set. We will show a comparison among the performances of these two settings.

A. Salinas dataset

A total of 120 independent runs of the cross validation were performed, corresponding to the different pairs that can be picked from the 16 classes.

Fig 6 shows the distribution of the mean accuracies of each cross validation. It can be seen that at least half of the pairs obtained a mean classification accuracy of 100%, and only one pair obtained a accuracy smaller than 85% (*Grapes_untrained* x *Vinyard_untrained*). Table IV shows the mean accuracies for all the pairs and for each variation. It can be seen that, for this dataset, none of the variations presented a significantly better performance than the other.

In Fig. 7, we plot the silhouette scores of all the best individuals obtained, versus the classification accuracy that



(a) (b)

Fig. 7. Plot of the silhouette score vs. accuracy for the Salinas dataset. Charts (a) and (b) correspond to indices learned without and with validation set, respectively.

they achieved, in order to evidence some kind of correlation. As in Fig. 6, we separated the two variations. It can be evidenced that a good silhouette score can lead to a good classification performance. The Pearson correlation coefficient was of approximately 0.7 for experiments with and without validation.

Fig. 8 shows two of the *non-trivial* indices learned that yielded a high classification accuracy (≥ 0.99). Note that for some classes, a single band was enough to achieve a good separation (e.g., *Soil_vinyard_develop* vs. *Vinyard_vertical_trellis* with band number 10). In order to visualize the indices, the generated images of the corresponding scenes (using min-max normalization) had to be equalized, due to pixels with extremely low or high values with respect to the rest. The equalized image was binarized with a manually-selected threshold since, even after equalization, the differences between the regions that were separated were not so evident in the intensities of the pixels. The equalization and binarization were necessary only for those indices that were complex.

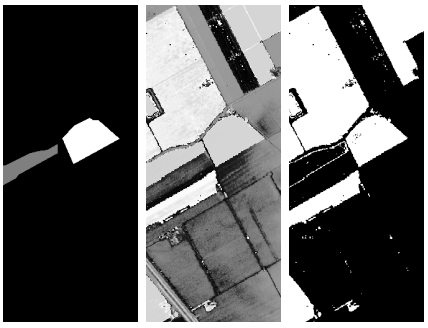
Fig 9 shows one of the indices learned that yielded poor results. It can be seen that not even by binarizing the image, it was possible to evidence some separation of the involved regions.

Fig. 10(a) shows the evolution curves of one of the experiments that best represented the mean behavior of the rest. Each generation returned two individuals: the one with the best silhouette without considering validation set (blue line) and the one with the best silhouette considering validation set (cyan and magenta lines). The score in the validation set was often slightly higher than the score in the training set, however, in general, the three curves were very close among them. This means that, for this dataset, the generalization capacity of the method is good, since no over or under-fitting is observed.

B. Thetford Mines dataset

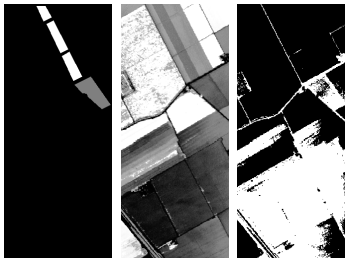
A total of 21 independent runs of the cross validation were performed, corresponding to the different pairs that can be picked from the 7 classes.

Fig 11 shows the distribution of the mean accuracies of each cross validation. For this data set, approximately half of



```
(b116 - b136) % (((b122 + b18) % ((b116 - b136) - ((((b116 - b40) - (b136 -
b41)) - b189) * (((b189 + b213) - ((((b116 - b40) - (b136 - b41)) - b189) *
(b122 + ((b189 + b213) - b122)))) + (((b122 + b18) - b41) - b122))))) * (b18
+ (((b122 + b37) % (((b116 - b40) - (b136 - b41)) - (b189 + (b189 + ((((b189
+ b213) - ((b41 - b189) * (b122 + ((b189 + b213) - b122)))) - b122) % ((b116
- b40) + b37)))))) * (b122 + (b122 + b37)))))
```

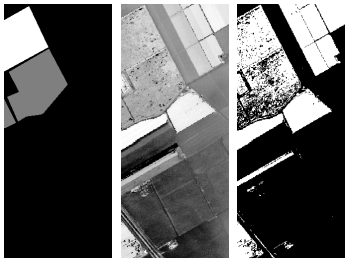
(a)



```
b133 - ((((((((((b200 % b133) % b133) * (b152 + b33)) % b133) % b133) *
((((((b200 % b133) % b133) * ((((((b200 % b133) % b133) * (b152 + b33)) %
b133) * ((b81 * (b30 + b197)) + (((((b200 % b133) % b133) * ((((b200 % b133)
% b133) * (b200 % b133)) * (b200 % b133))) % b133) % ((b33 % b133) * (b30 +
b197))))) * (b200 % b133))) % b133) % b133) * b33)) * b210) * (b152 + b33)) *
b210) * (b212 * b191))
```

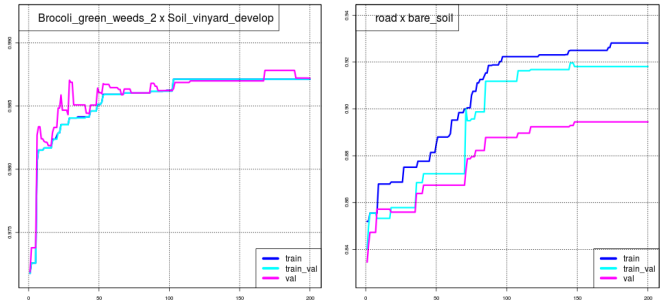
(b)

Fig. 8. Visualisation of the spectral indices obtained to separate *Brocoli_green_weeds_1* vs. *Brocoli_green_weeds_2* (a) and *Fallow* and *Fallow_smooth* (b) from the Salinas dataset. From left to right: ground-truth of the classes to be separated, equalized image of the spectral index and binarization of the equalized image.



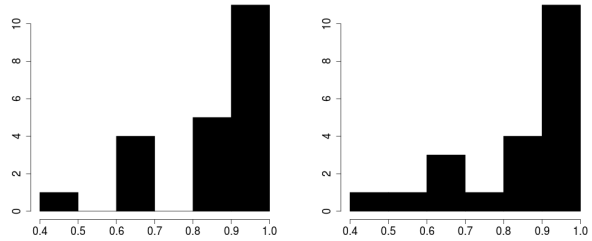
```
(((b60 - b101) % (b207 * b40)) % (b201 * ((b204 + b31) % b31))) % (b201 *
(((b207 * b40) % ((b60 - b101) % ((b61 % b70) % b70))) % (b201 + ((((((b61 %
b70) % ((b61 % b70) % b70)) * (b61 % ((b61 % b70) * b132))) * b31) - b31) %
((b140 % b32) % (b76 + b132))))))
```

Fig. 9. Visualisation of the spectral index obtained to separate *Grapes_untrained* vs. *Vineyard_untrained* from the Salinas dataset. From left to right: ground-truth of the classes to be separated, equalized image of the spectral index and binarization of the equalized image.



(a) (b)

Fig. 10. Silhouette curves of the GP through the generations for the Salinas (a) and Thetford Mines (b) data sets. The blue curve represents the score on training data of the best individual found without considering validation set. The cyan curve represents the score on training data of the best individual found considering validation set. The magenta curve represents the score on validation data of the same individual of the cyan curve.



(a) (b)

Fig. 11. Histograms of the classification mean accuracies for the 21 pairs of Thetford Mines' classes. Charts (a) and (b) show the distribution for indices learned without and with validation set, respectively.

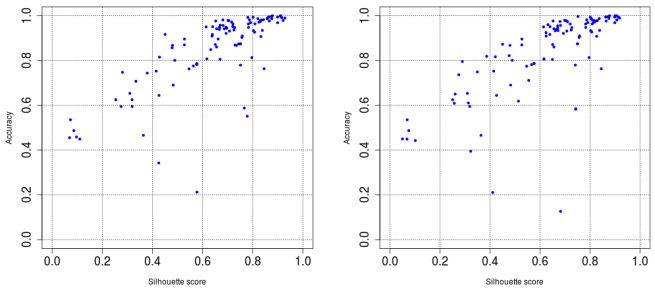
TABLE V
OVERALL MEAN ACCURACIES FOR THE THETFORD MINES DATASET

	Without validation	**With validation**
Mean accuracy	0.8529547	0.8398459

the pairs obtained a mean accuracy of 100% and just a few obtained mean accuracies between 80% and 100%. There were pairs with a very poor performance, like *reed_rood* x *grey_roof* that obtained accuracies below 50%. Table V shows the mean accuracies for all the pairs and for each variation. Again, the impact of the use of a validation set to re-rank the individuals was not evident.

In Fig. 12, we plot the silhouette score of all the best individuals obtained, versus the classification accuracy that they achieved, in order to evidence some kind of correlation. As in Fig. 11, we separated the variations. It can be evidenced that a good silhouette score can lead to a good classification performance for this dataset too. The Pearson correlation coefficient was of approximately 0.85 for experiments with and without validation.

Fig. 13 shows two of the indices learned that yielded a high classification accuracy. Unlike Salinas dataset, there were

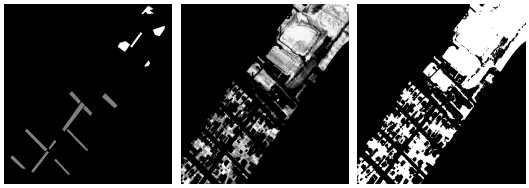


(a) (b)

Fig. 12. Plot of the silhouette score vs. accuracy for the Thetford Mines dataset. Charts (a) and (b) correspond to indices learned without and with validation set, respectively.

```
(((b0 - b41) - (b22 % b60)) % (((b14 + b10) * (b10 * (b27 * (b3 * b3)))) *
(b10 * (b3 * b3)))) % ((b10 * (b3 * b3)) % (b3 + ((((b8 + b29) - b71) % ((b8
+ b29) * ((b14 + b75) * (b8 + b29)))) % (((b14 + b10) * (b3 * (b44 + b55))) %
((((b71 % (b14 + b10)) % ((b0 - b41) * ((b8 + b29) * (b10 * (b27 * (b3 *
b3)))))) % (b10 * (b3 * b3))) + b55)))))
```

(a)



```
((b61 + b17) - ((((b74 - (b18 + b6)) % (((b11 * b43) * (((b11 * b43) *
((((b11 * b43) * ((b11 * b43) * ((b53 + b5) * ((b53 + b5) * b61)))) * b61) *
((b6 * ((b6 * ((b11 * b43) * (b53 + b5))) * (b18 + b6))) * b61))) * b43)) *
(b48 + b15))) % b30) - (b48 + b15))) % ((((b11 * b43) * (b53 + b5)) * ((b6 *
((b53 + b5) * ((b53 + b5) * b61))) * b61)) * ((b11 * b43) * (b61 + b17)))
```

(b)

Fig. 13. Visualisation of the spectral indices obtained to separate *road* vs. *trees* (a) and *road* and *bare soil* (b) from Thetford Mines dataset. From left to right: ground-truth of the classes to be separated, equalized image of the spectral index and binarization of the equalized image.

no indexes involving single classes. Resulting images were equalized and binarized as well. The scenes were cropped in order to show with more detail the separation.

Fig 14 shows one of the indices learned that yielded poor results. As it can be observed, likewise the one shown for Salinas dataset in Fig. 9, no separation can be evidenced.

Fig. 10(b) shows the evolution curves of one of the experiments that best represented the mean behavior of the rest. In general, the score in the validation set was lower than the score

```
((b41 * (b68 * (b41 * b68))) * (b41 * b41)) % (((b28 - b51) % b81) % (b41 %
(((((b28 - b51) * ((((((((((b28 - ((b28 - b51) % b41)) % (b41 * b41)) * ((b41
 * b41) % (b41 * b41))) + b49) + b49) % b68) % b41) % b41) % (b41 * b41)) %
 ((b41 * (b41 % b41)) * b41))) + b49) % (b41 * b41)) % ((b41 * b41) % b41))))
```

Fig. 14. Visualisation of the spectral index obtained to separate *red roof* vs. *grey roof* from Thetford Mines dataset. From left to right: ground-truth of the classes to be separated, equalized image of the spectral index and binarization of the equalized image.

in the training set. However, as it can be seen in Fig. 12, for high silhouette scores, variations in these scores yield much smaller variations in the classification accuracy. Therefore, in terms of performance, it can be said that although the validation curve is under the training curves, the generalization capacity of the method is still high.

VII. Conclusions and future work

This paper has introduced a Genetic Programming framework for hyper- and multi-spectral band selection and combination and described its use in binary classification tasks. The proposed method has as the main objective to determine along evolutionary iterations the band combination function that best separates the pixel value distributions observed for different classes. Performed experiments considering two different data sets demonstrated that the proposed approach is very effective in classifying remote sensing image pixels given the band combination function found by the GP framework.

Future work will focus on the extension of the proposed method to handle multi-class classification problems by means of ensemble methods, the investigation of novel fitness functions based on distribution-based similarity measures, the use of other classification methods (e.g., probability-based), and other arithmetic operators in the GP tree representation.

Acknowledgments

The authors would like to thank CAPES, CNPq, FAPEMIG, and the Microsoft-FAPESP virtual institute (grants #2013/50155-0 and #2013/50169-1) for the financial support.

References

[1] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, pp. 138–142, Jan. 2011.

[2] A. Ifarraguerri and M. Prairie, "Visual method for spectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, pp. 101–106, Apr. 2004.

[3] R. Huang and M. He, "Band selection based on feature weighting for classification of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, pp. 156–159, Apr. 2005.

[4] R. Nakamura, L. Garcia Fonseca, J. A. dos Santos, R. da S. Torres, X.-S. Yang, and J. P. Papa, "Nature-inspired framework for hyperspectral band selection," *IEEE TGRS*, vol. 52, pp. 2126–2137, Apr. 2014.

[5] G. Zhu, Y. Huang, J. Lei, Z. Bi, and F. Xu, "Unsupervised hyperspectral band selection by dominant set extraction," *IEEE TGRS*, vol. 54, no. 1, pp. 227–239, 2016.

[6] S. Jia, Y. Qian, and L. Shen, "Unsupervised band selection for hyperspectral imagery classification without manual band removal," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.*, vol. 5, no. 2, pp. 531–543, Apr. 2012.

[7] R. Hedjam and M. Cheriet, "Hyperspectral band selection based on graph clustering," in *ISSPA*, Jul. 2012, pp. 813 –817.

[8] G. C. T. Jee Cheng Wu, "Unsupervised cluster-based band selection for hyperspectral image classification," in *ICACSEI*, Jul. 2013.

[9] B. Guo, S. Gunn, R. I. Damper, and J. D. B. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, pp. 522–526, Oct. 2006.

[10] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE TGRS*, vol. 54, no. 1, pp. 88–102, 2016.

[11] P. Groves and P. Bajcsy, "Methodology for hyperspectral band and classification model selection," in *WARSD*, Oct. 2003, pp. 120–128.

[12] P. Bajcsy and P. Groves, "Methodology for hyperspectral band selection," *Photogrammetric Engineering and Remote Sensing Journal*, vol. 70, pp. 793–802, 2004.

[13] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.

[14] L. C. B. dos Santos, S. J. F. Guimaraes, and J. A. dos Santos, "Efficient unsupervised band selection through spectral rhythms," *IEEE Journal of Sel. Top. in Sig. Proc.*, vol. 9, no. 6, pp. 1016–1025, Sept 2015.

[15] S. Patra, P. Modi, and L. Bruzzone, "Hyperspectral band selection based on rough set," *IEEE TGRS*, vol. 53, no. 10, pp. 5495–5503, 2015.

[16] J. Kittler, "Feature set search algorithm," in *Pattern Recognition and Signal Processing*, 1978, pp. 41–60.

[17] X. Geng, K. Sun, L. Ji, and Y. Zhao, "A fast volume-gradient-based band selection method for hyperspectral image," *IEEE TGRS*, vol. 52, no. 11, pp. 7111–7119, Nov 2014.

[18] Y. Yuan, G. Zhu, and Q. Wang, "Hyperspectral band selection by multitask sparsity pursuit," *IEEE TGRS*, vol. 53, no. 2, pp. 631–644, Feb 2015.

[19] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA, USA: MIT Press, 1992.

[20] C. W. B. Fonlupt and D. Robilliard, "Genetic Programming with Dynamic Fitness for a Remote Sensing Application," *PPSN*, pp. 191–200, 2000.

[21] C. Chion, J.-a. Landry, and L. Da Costa, "A genetic-programming-based method for hyperspectral data information extraction: Agricultural applications," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 8, pp. 2446–2457, 2008.

[22] C. Puente, G. Olague, S. V. Smith, S. H. Bullock, A. Hinojosa-Corona, and M. A. González-Botello, "A genetic programming approach to estimate vegetation cover in the context of soil erosion assessment," *Photogrammetric Engineering & Remote Sensing*, vol. 77, no. 4, pp. 363–376, 2011.

[23] B. J. Ross, A. G. Gualtieri, F. Fueten, and P. Budkewitsch, "Hyperspectral image analysis using genetic programming," *Applied Soft Computing*, vol. 5, no. 2, pp. 147–156, 2005.

[24] P. J. Rauss, J. M. Daida, and S. Chaudhary, "Classification of spectral imagery using genetic programming," *Ann Arbor*, vol. 1001, p. 48109, 2000.

[25] S. J. Perkins, J. P. Theiler, S. P. Brumby, N. R. Harvey, R. B. Porter, J. J. Szymanski, and J. J. Bloch, "Genie: a hybrid genetic algorithm for feature classification in multispectral images," *Proc. SPIE*, vol. 4120, pp. 52–62, 2000.

[26] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53 – 65, 1987.

[27] R. da S. Torres, A. X. Falcão, M. A. Gonçalves, J. P. Papa, B. Zhang, W. Fan, and E. A. Fox, "A genetic programming framework for content-based image retrieval," *Pattern Recogn.*, vol. 42, no. 2, pp. 283–292, Feb. 2009.