Title: A Study on the Effectiveness of Different Independent Component Analysis Algorithms for Hyperspectral Image Classification

This paper appears in: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

Date of Publication: July 2014

Author(s): Nicola Falco, Jon Atli Benediktsson, Lorenzo Bruzzone

Volume:

Issue:

Page(s):

DOI: 10.1109/JSTARS.2014.2329792

A Study on the Effectiveness of Different Independent Component Analysis Algorithms for Hyperspectral Image Classification.

Nicola Falco, *Student Member, IEEE,* Jón Atli Benediktsson, *Fellow, IEEE,* and Lorenzo Bruzzone, *Fellow, IEEE*

***Abstract*—This paper presents a thorough study on the performances of different Independent Component Analysis (ICA) algorithms for the extraction of class-discriminant information in remote sensing hyperspectral image classification. The study considers the three implementations of ICA that are most widely used in signal processing, namely Infomax, FastICA and JADE. The analysis aims to address a number of important issues regarding the use of ICA in the RS domain. Three scenarios are considered and the performances of the ICA algorithms are evaluated and compared against each other, in order to reach the final goal of identifying the most suitable approach to the analysis of hyperspectral images in supervised classification. Different feature extraction and selection techniques are used for dimensionality reduction with ICA and are then compared to the commonly used strategy, which is based on pre-processing data with Principal Components Analysis (PCA) prior to classification. Experimental results obtained on three real hyperspectral data sets from each of the considered algorithms are presented and analyzed in terms of both classification accuracies and computational time.**

***Index Terms*—Independent component analysis (ICA), hyperspectral images, feature extraction, dimensionality reduction (DR), supervised classification, remote sensing.**

I. Introduction

HYPERSPECTRAL images are an important information source for analysing and understanding the processes that occur on the surface of our planet. Due to their capabilities in providing highly detailed representation of the spectral signature of the different materials present in the scene, the processing of hyperspectral images can provide results that are suitable for a wide range of applications. In the Earth Observation domain, one of the major challenges is the classification of remote sensing images, which is the process of identifying the diverse classes of coverage of an investigated area. In particular, the analysis of hyperspectral images is a challenging and computationally expensive task as these data are characterized by hundreds of narrow spectral channels offering a huge quantity of often redundant spectral information. The large number of spectral channels may result in the so called "curse of dimensionality" (Hughes' phenomenon) [1], which arises when the number of spectral features is high and only relatively few training samples are available, affecting the generalization capability of the classifier. To overcome this problem, the common strategy is to select or extract a subset of features from the original hyperspectral data set. The resulting representation contains the useful information while the redundant and the noisy components are discarded after the dimensionality reduction (DR) process. Different feature subsets can be obtained by exploiting various feature selection/extraction techniques in accordance to the task to be accomplished. For example, by exploiting the Karhunen-Loève transform [2] (also known as Principal Component Analysis (PCA)), it is possible to obtain a reduced subset while achieving excellent data compression [3] and a good representation in terms of minimum mean square error. Alternatively, the Maximum Noise Fraction approach [4] provides a subset whose components are ordered based on the noise fraction contained in the features. Regarding the classification task, where the class separability is the variable that has to be considered, techniques based on discriminant analysis [2] are more adequate since they can exploit the available prior information. Important techniques based on discriminant analysis are Discriminant Analysis Feature Extraction (DAFE) [2], Decision Boundaries Feature Extraction (DBFE) [5], projection pursuit (PP) [6] and Non-parametric Weighted Feature Extraction (NWFE) [7]. The problem of blind source separation appears in various signal processing fields such as biomedical signal analysis where Independent Component Analysis (ICA) [8] is proven to be an efficient method for the interpretation of signals, e.g., in electroencephalography (EEG), in electrocardiography (ECG), in electromyography (EMG), in magnetoencephalography (MEG) and in electronystagmography (ENG) [9]–[13]. ICA is a well known unsupervised blind source separation technique, extensively used in several fields, aimed at finding statistically independent components (ICs) by only considering the observation of mixture signals. ICA-based methods are also applied to geophysical data processing, data mining, speech enhancement, image recognition and wireless communications [14]. Recently, ICA has received attention in the hyperspectral remote sensing domain and in particular for feature dimensionality reduction [15] and unmixing [16], [17], while applications of ICA for classification have also been exploited [18]–[23]. In hyperspectral remote sensing, ICA extracts the source components that generate the mixed signal measured by the sensor and the independent components refer to the different classes presented

N. Falco is with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland, and also with the Department of Information Engineering and Computer Science, University of Trento, 38050 Trento, Italy (e-mail: nicolafalco@ieee.org).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland (e-mail: benedikt@hi.is).

L. Bruzzone are with the Department of Information Engineering and Computer Science, University of Trento, 38050 Trento, Italy (e-mails: lorenzo.bruzzone@ing.unitn.it).

in the scene. Several algorithms have been proposed in the literature for implementing ICA based on the maximization of different criteria. The different algorithms provide diverse feature sets for classification. However, only a limited number of studies addresses the comparative performance of these algorithms. The available studies are in most cases related to biomedical signals analysis [24]–[27] yielding results that are not consistent in terms of the most efficient ICA algorithm. All the review papers in the aforementioned domain agree on the identification of the three most prominent algorithms, that are Infomax [28], FastICA [29] and JADE [30]. However, an in-depth comparative study that addresses simultaneously fundamental questions on the properties and the efficiency of ICA implementations for the analysis of hyperspectral remote sensing images is still missing.

In the literature, a common approach is to apply ICA after DR, which is usually carried out by PCA. This approach is applied in [18], [19], where PCA is performed firstly and then the ICA is applied to the most important principal components with the accumulative variance of 99% and 98.58%, while the remaining components are discarded. In other studies [21]–[23], JADE and FastICA are used to extract subsets of ICs by exploiting the PCA phase implemented in the algorithms for dimensionality reduction. As mentioned before, PCA aims to globally decorrelate the data and maximize the variance. The main limitation of PCA is that it is based on using the global second order statistics for the whole image. Consequently, the sensitivity to critical classes composed of a small number of pixels is reduced [31]. It is also well known that the criterion for retaining a certain number of components based on the calculation of the accumulated sum of eigenvalues is not an effective measure in terms of class discriminant, as demonstrated in [32]. Thus, PCA should not be used as a pre-processing tool for classification purposes [33]. Note that, according to the studies conducted in [8], ICA results obtained after PCA are in general not sufficient to estimate the ICs, since after the use of PCA only information on a subset of orthogonal components is available. In general, some weak ICs may be hidden in the dimensionality reduction process. An attempt to identify a better pre-processing approach than PCA is performed in [20], where a Noise-adjusted Principal Components (NAPC) is used for dimensionality reduction. The obtained results show that the principal components from the NAPC can better maintain the object information in the original data than those from PCA, allowing the ICA to provide better object classification.

The aim of this work is twofold; firstly, we aim at identifying an effective strategy for the extraction of class-discriminant features with ICA. In the analysis we consider different supervised feature extraction and selection approaches to dimensionality reduction (DR), which are investigated as pre-processing before applying ICA. Secondly, we aim at addressing the lack in the literature of an extended comparative study on the three most frequently used implementations of ICA in the broader field of signal processing: Infomax, FastICA and JADE, aiming at assessing the most efficient and reliable methodology to follow when employing the ICA technique for accurate and cost efficient classification of hyperspectral images. Importantly the computational cost is assessed in relation to the number of samples used for the source estimation.

The paper is organized as follows. In Section II, the general ICA problem is introduced, and the three different algorithms: Infomax, FastICA and JADE are briefly presented. Section III describes the experimental setup, explaining the strategy adopted for each experiment. Section IV contains the experimental results. Finally, conclusions and future steps are drawn in Section V.

II. Independent Component Analysis

Let us consider n mixtures of random variables $x_1, x_2, ..., x_n$ which are defined as a linear combination of n random variables $s_1, s_2, ..., s_n$. The mixing model can be written as:

$$x_i = a_{i,1}s_1 + a_{i,2}s_2 + ... + a_{i,n}s_n \quad i = 1, ..., n. \tag{1}$$

In terms of random vectors, the model can be rewritten as:

$$\mathbf{x = As}, \tag{2}$$

where $\mathbf{x} = [x_1, x_2, ..., x_n]^T$ is the observed vector, $\mathbf{A}$ is the unknown mixing matrix with element $a_{ij}, i, j = 1, ..., n$ (which are real coefficients) and $\mathbf{s} = [s_1, s_2, ..., s_n]^T$ is the unknown source vector. By estimating the unmixing matrix of $\mathbf{A}$, called $\mathbf{W}$, the $\mathbf{s}$ vector that represents the independent components (ICs) is obtained by:

$$\mathbf{s = Wx}. \tag{3}$$

The estimation of the ICA model is possible if the following assumptions and restrictions are satisfied: 1) the sources are statistically independent; 2) the independent components must have a non-Gaussian distribution; 3) the unknown mixing matrix $\mathbf{A}$ is assumed square and full rank. Under these conditions, the ICA model can be rewritten as:

$$\mathbf{y} = \mathbf{Wx} \simeq \mathbf{s}, \tag{4}$$

where $\mathbf{W} \simeq \mathbf{A}^{-1}$. The problem can be solved by estimating $\mathbf{W}$ to obtain $\mathbf{y}$ that represents the best possible approximation of $\mathbf{s}$. Nevertheless, since $\mathbf{W}$ and $\mathbf{s}$ are unknown in the ICA model, three ambiguities necessarily hold:

1) The variances (energies) of the independent components cannot be determined. That is because any scalar multiplier in one of the sources s_i could always be canceled by dividing the corresponding column $\mathbf{a}_i$ of $\mathbf{A}$ by the same scalar.
2) For similar reasons, also the order of the independent components cannot be ranked.
3) The sign cannot be determined. This means that dark and bright regions may have the same meaning, which is not critical in most applications.

In this paper, three different implementations of ICA are investigated for feature extraction. In particular, the analysis focus on the Infomax, FastICA and the JADE algorithms, which are briefly introduced in the next subsection. As mentioned previously, the scope of this study is to present a complete comparison among the most widely used ICA algorithms in the remote sensing field. For the sake of scientific concreteness

we attempted to exploit more recent implementations of ICA that are used in the broader signal processing field. To the best of our knowledge, one of the most recent implementation of ICA stated to outperform FastICA is RobustICA [34]. This method is presented in the next section. However, since the computational cost was excessively high, the method is evaluated in only one experiment and the results are discussed in the corresponding section.

An important issue that characterizes ICA transformation is the non-prioritization of the ICs. Accordingly, multiple ICA applications result in different IC sets, which are diverse both in the order of appearance and in the content, thus making a performance comparison inconsistent. This behaviour is caused by the fact that ICA uses random vectors as initial projections. Wang and Chang addressed this problem in [15] proposing an initialization algorithm in conjunction with the virtual dimensionality (VD) [32] to generate an appropriate set of initial projections. The algorithm was designed for FastICA. However, in order to exploit the original setup without modifying the algorithms, the identity matrix of size $n \times n$ has been chosen as a common initialization for the ICA transformation. It is possible that in some cases the identity matrix gives worse results than a random initialization in terms of convergence time. The advantage in using a constant initialization is the consistency of the obtained components and their ordering.

A. Infomax Algorithm

Infomax [28] is based on the minimization of the mutual information between the input and output of a neural network with nonlinear units. The mutual information of a pair of random variables x and y can be defined as:

$$I(x;y) = H(x) - H(x \mid y) \tag{5}$$

where $H(x \mid y)$ is the conditional entropy defined as:

$$H(x \mid y) = H(x,y) - H(y). \tag{6}$$

Considering the entropy as a measurement of uncertainty and the mutual information as a measurement of the dependency between random variables, the matrix $\mathbf{W}$ is determined so that the mutual information among the components of the transformed vector $\mathbf{y}_i$ is minimized. The convergence is quite slow since the inverse matrix has to be computed at each iteration.

The algorithm's implementation used in this work is a part of the *EEGLAB* toolbox [9], [35]. The algorithm performs ICA decomposition using the logistic infomax ICA algorithm developed in [28] with a natural gradient feature as defined by Amary, Cichocki and Yang [36]. The algorithm performs a sphering (whitening) of the data in order to increase the convergence rate. This means that the unmixing matrix that is processed becomes $\mathbf{W} = weights\ matrix * sphere\ matrix$.

B. FastICA Algorithm

The FastICA algorithm proposed in [8] is a very efficient and robust method for ICA. It exploits the negentropy J, which is a measurement of non-Gaussianity that gives a measure of the distance from normality. It is defined as:

$$J(\mathbf{y}) = H(\mathbf{y}_{Gaussian}) - H(\mathbf{y}) \tag{7}$$

with $\mathbf{y}$ being a random vector, $H(\mathbf{y})$ the entropy of $\mathbf{y}$ and $H(\mathbf{y}_{Gaussian})$ the entropy of a Gaussian random vector with the covariance matrix equal to the one of $\mathbf{y}$. Negentropy is always nonnegative and is zero only in case of Gaussian distribution. Because of the complexity of (7), the following moment-based approximation has been introduced [8]:

$$J(y) \propto [E\{(G(y)\} - E\{G(v)\}]^2 \tag{8}$$

where y is a standardized non-Gaussian variable, v is a standardized Gaussian variable and G is a non-quadratic function. The learning rule for FastICA is based on a fixed-point iteration scheme [29] that has been found to be considerably faster than using gradient descent methods for solving ICA. Before the FastICA algorithm can be applied, the input vector data should be centered and whitened. The scheme finds the maximum of the non-Gaussianity of $\mathbf{w}^T\mathbf{x}$. The basic fixed-point iteration for the estimation and decorrelation of one single independent component is:

$$\begin{aligned} \mathbf{w}_{i+1} &\leftarrow E\{\mathbf{x}g(\mathbf{w}_i^T\mathbf{x})\} - E\{\acute{g}(\mathbf{w}_i^T\mathbf{x})\}\mathbf{w}_i \\ \mathbf{w}_{i+1} &\leftarrow \mathbf{w}_{i+1} - \sum_{j=1}^{i}(\mathbf{w}_{i+1}^T\mathbf{w}_j)\mathbf{w}_j, \end{aligned} \tag{9}$$

where $g(u)$ is a non-quadratic function that represents the derivative of the non-quadratic function G in (8). The algorithm converges when the old and new values of $\mathbf{w}$ (where $\mathbf{w}$ represents one row of $\mathbf{W}$), point in the same direction. The FastICA algorithm can be used to perform projection pursuit as well, thus providing a general-purpose data analysis method that can be used both in an exploratory fashion and for the estimation of independent components (or sources).

In this work, the broadly used FastICA package (version 2.5, 2005) has been used. The algorithm can estimate the ICs in two different ways: 1) deflationary orthogonalization, which is shown in (9), 2) symmetric orthogonalization, which is shown in (10). The first approach performs orthogonalization using the Gram-Schmidt method, estimating the ICs one by one, while the second approach estimates all the ICs in parallel. In our experiments, the second approach is used mainly for two reasons: 1) to avoid the cumulative error in the estimation, and 2) to estimate the ICs by a parallel computation, thus making the algorithm faster. In this case the basic fixed-point iteration in FastICA with symmetric orthogonalization is as follows:

$$\begin{aligned} \mathbf{w}_{i+1} &\leftarrow E\{\mathbf{x}g(\mathbf{w}_i^T\mathbf{x})\} - E\{\acute{g}(\mathbf{w}_i^T\mathbf{x})\}\mathbf{w}_i \\ \mathbf{W} &\leftarrow (\mathbf{W}\mathbf{W}^T)^{-\frac{1}{2}}\mathbf{W}\ with\ \mathbf{W} = (\mathbf{w}_1, \cdots, \mathbf{w}_m)^T. \end{aligned} \tag{10}$$

C. Joint Approximate Diagonalization of Eigenmatrices Algorithm

The Joint Approximate Diagonalization of Eigenmatrices (JADE) is a widely used and parameter-free implementation of ICA. In the pre-processing, a whitening transformation is performed on the mixtures, which makes the original components uncorrelated and thus independent in terms of second

order statistics, and the unmixing matrix $\mathbf{W}$ orthogonal. The approach exploits the concept of cumulant tensor, which can be seen as a generalization of the covariance matrix. Let us consider the whitened unmixing matrix $\mathbf{W}$ and the cumulant tensor $\mathbf{F(M)}$, which is a linear symmetric operator. We can define an eigenmatrix $\mathbf{M}$ such that

$$\mathbf{F(M)} = \lambda \mathbf{M} \tag{11}$$

where every eigenmatrix has the form $\mathbf{M} = \mathbf{w}_n \mathbf{w}_n^T$, where $\mathbf{w}_n$ is a row of the unmixing matrix $\mathbf{W}$. Thus, knowing the eigenmatrix of the tensor, it is easy to obtain the independent components. The main problem is that the eigenvalues are not distinct, and thus, the matrices cannot be uniquely defined. Considering that $\mathbf{F}$ is a linear combination in the form $\mathbf{w}_n \mathbf{w}_n^T$, it can be observed that the matrix $\mathbf{W}$ diagonalizes $\mathbf{F(M)}$ for any $\mathbf{M}$. This means that it is important to choose a set of n different matrices $\mathbf{M}_i$ that makes the matrices $\mathbf{WF(M}_i\mathbf{)W}^T$ as diagonal as possible. The diagonality can be measured as the sum of squares of diagonal elements and is defined as:

$$J_{JADE}(\mathbf{W}) = \sum_i \|diag(\mathbf{WF(M}_i\mathbf{)W}^T)\|^2 \tag{12}$$

One method of join approximate diagonalization of the $\mathbf{F(M}_i\mathbf{)}$ is to maximize J_{JADE}. The algorithm used in the experiment analysis was developed in [30].

D. RobustICA

RobustICA [34] is a recent method for deflationary ICA, in which the kurtosis is the general contrast function to be optimized. The method performs the optimization by a computationally efficient technique based on an optimal step size (adaption coefficient). The technique computes algebraically (i.e., without iterations) the step size globally optimizing the kurtosis in the search direction at each extracting vector update. In the derivation of the algorithm, no-simplifying assumptions concerning specific type of sources (real or complex, circular or noncircular, sub-Gaussian or super-Gaussian) are involved. The method presents a number of advantages with significant practical impact when compared to other kurtosis-based algorithms such as the original FastICA and its variants:

- Prewhitening is not required, so that the performance limitations it imposes can be avoided and the sequential extraction (deflation) can be carried out, e.g., via linear regression.
- Sub-Gaussian or super-Gaussian sources can be extracted in the order specified by the user if the Gaussianity character of the sources is known in advance.
- The optimal step-size technique provides some robustness to the presence of saddle points and spurious local extrema in the contrast function.
- In the experimental analysis performed in [34], the method shows a very low computational cost measured in terms of source extraction quality versus number of operations, even without prewhitening.

For further details about the implementation, it is suggested referring to the paper [34].

III. Experimental Setup

A. Design of Experiments and Investigations

The analysis presented in this paper aims at identifying which ICA implementation provides better results in terms of classification accuracy and computational cost. This is studied in three scenarios:

1) *Low-dimensional space:* This represents the most common scenario in remote sensing image analysis, where ICA is exploited. In general a small subset of features is obtained by performing dimensionality reduction on a high-dimensional feature space. The ICs are then extracted by processing the reduced subset. In the analysis we consider the use of a number of feature extraction and feature selection methods used for dimensionality reduction. The obtained results are compared against the general case in which PCA is exploited for feature reduction. The goal is to analyze and compare the performance of the three ICA algorithms applied to different subsets of features, identifying which pair of ICA algorithm and feature reduction technique gives the best classification accuracy.
2) *High-dimensional space:* The performance of ICA is evaluated by considering the entire data set. The obtained feature space is then reduced by selecting the most informative features by exploiting a supervised feature selection algorithm. These features are then used in classification. The aim is to investigate the effectiveness of the ICA algorithms in extracting useful independent components directly from the original feature space, without initially projecting the data into a smaller subspace.
3) *Spatial downsampling:* In this scenario the ICA is applied to subsets of image samples obtained by spatially downsampling the original image. The goal is to investigate how the performance of the ICA is affected by decreasing the number of samples used for the source estimation, and thus if it is possible to achieve classification accuracies that are similar to those obtained by using the entire data set. The exploitation of a reduced spatial subset would also positively affect the computational time of the ICA.

In the analysis, based on the above scenarios three experiments are designed as described below.

1) Experiment I - Low-dimensional space: As discussed in Section I, hyperspectral images are usually pre-processed by reducing the feature space in order to decrease the computational cost, discard redundant information and mitigate the noise contribution. Regarding the use of ICA for feature reduction, the most common strategy in remote sensing image analysis is to apply the PCA technique to the original image followed by the ICA. PCA is used to extract the high-variance components while filtering out the low-variance components. It is worth noting that the use of PCA is encouraged by the fact that it is implemented in the ICA as part of the algorithm for whitening purposes (see Section II), where the user can decide to perform the dimensionality reduction by choosing the number of components to be retained. However,

the PCA transformation provides a subset of components that after selection does not preserve class-separability. This also affects the independent components. In this experiment a different strategy is proposed and investigated. The aim is to provide a reduced feature set where the class-information is preserved and used as input to the ICA, avoiding the use of the PCA-based reduction approach. To this purpose, considering the context of supervised classification, the dimensionality reduction is performed by exploiting three supervised feature selection and extraction techniques, namely the Steepest Ascent (SA) search algorithm (in which the Jeffries-Matusita distance is used as the criterion function in feature selection), the Local Fisher Discriminant Analysis (LFDA), and the Non-parametric Weighted Feature Extraction (NWFE). The strategy adopted in the experiment consists of three steps: a) dimensionality reduction; b) application of the ICA to the obtained feature subset; c) evaluation in terms of classification accuracy of the effectiveness of the extracted ICs in discriminating the classes. The procedure is repeated for every ICA algorithm, considering different subsets of the retained components, starting from a minimum of 5 components up to 40 components. In the paper, the different strategies are referred as *DR-approach-ICA*, where *DR-approach* is one of the feature extraction/selection techniques aforementioned (e.g., in the case of NWFE the strategy would be NWFE-ICA). The background information on the feature extraction and feature selection approaches that are used in this work is provided in Section III-C.

2) Experiment II - High-dimensional space: Experiment II aims at investigating the effectiveness of the independent components obtained by considering the entire original hyperspectral data set, without performing any feature reduction (which reduces both redundancy and noise but may introduce information loss). The strategy adopted in the experiment is defined as follows: a) ICA is applied to the entire data set and all the components are retained; b) the most informative components are selected by applying the SA feature selection algorithm; c) the effectiveness of the subset is assessed in terms of classification accuracy. Also, in this case we take advantage of the training samples in order to select the best independent components. As already mentioned in Section III-A1, JADE's computation load is extremely high when the dimensionality of the feature space becomes large. This is due to the fact the JADE implementation has to estimate the initial vector of n eigenmatrices whose dimension is $n \times n$ (see Section II-C), where n is the number of the sources to estimate. When n increases, the size of the initial projection increases as the cube of n, requiring the availability of a significant quantity of physical memory. For these reasons, JADE is not used in this experiment and in general it should be avoided when a high-dimensional space is considered.

3) Experiment III - Spatial downsampling: The third experiment aims at investigating the effectiveness of the ICA in extracting informative components when applied to a downsampled data set (i.e., only a portion of the total number of pixels is analyzed). The analysis consists of seven sub-experiments. In the first three sub-experiments the sampling rate is decreased by three different integer factors: 2, 3, 4. In the last four sub-experiment, different sizes of training samples are considered. The experiment has been conducted considering both scenarios 1 and 2, i.e., low-dimensional space and high-dimensional space, respectively. However, the results obtained from the analysis of the Botswana and Hekla data sets in high-dimensional space are very poor, especially when Infomax is used. Thus, for this scenario, only the results obtained by using FastICA performed on the Salinas data set are reported.

B. ICA Parameter Tuning

In the experimental analysis, an implementation of each ICA algorithm based on MATLAB (The MathWorks, Inc.) scripting language is used.

1) Infomax: as mentioned in the Section II, the initial weight matrix is initialized as an identity matrix. The training stops when the weight-change goes below the predefined threshold value, which is set by default at 10^{-6} when $n < 33$ and 10^{-7} otherwise, or after a maximum number of ICA training steps of 512.

2) FastICA: different parameters need to be tuned. The non-quadratic function $g(u)$ is set as $\tanh(au)$ with $a = 1$, which is proven a good approximation of negentropy [8]. In order to avoid a random initialization, an identity matrix of size $n \times n$ is given in input as initial guess. In this work, the symmetric ortogonalization is chosen for the reasons explained in Subsection II-B. The algorithm stops when the convergence is reached, meaning that the weight-change is less than 10^{-4}, or when the maximum number of iterations, which is set at 1000, is reached.

3) JADE: this technique is parameter free, i.e., no tuning is needed. The only experimental parameter that can be tuned is related to the stopping criterion, which is thresholded at 10^{-6} by default.

4) RobustICA: the method requires the tuning of few parameters. Two different approaches of deflation are possible: 1) via ortogonalization, 2) via linear regression. In this work deflationary ortogonalization is used. The threshold for statistical-significant termination test is set at 10^{-4}, while 1000 is the maximum number of possible iterations for each extracted source.

C. Feature Reduction

This section provides a briefly introduction to the feature reduction techniques used in this work.

1) Steepest Ascent (SA) Feature Selection: The supervised feature selection is based on the sub-optimal Steepest Ascent search algorithm using as criterion function the Jeffries-Matusita distance. The strategy is based on the search for constrained local extremes in a discrete binary space. More information can be found in [37].

2) Local Fisher Discriminant Analysis (LFDA): It is a linear supervised dimensionality reduction method. It combines the ideas of Fisher's Discriminant Analysis (FDA) [38] and Locality Preserving Projections (LPP) [39]: between-class separability is maximized while within-class local structure is preserved. LFDA has an analytic form of the embedding

matrix and the solution can be easily computed just by solving a generalized eigenvalue problem. Therefore, LFDA is scalable to large data sets and computationally reliable. More information can be found in [40].

3) Nonparametric Weighted Feature Extraction (NWFE): the NWFE algorithm [7] takes advantage of the desirable characteristics of DAFE and Nonparametric Discriminant analysis (NDA) [41], while avoiding their shortcomings. DAFE is fast and easy to apply, but it is able to extract only $L-1$ features, with L the number of classes. This limitation reduces the performance particularly when the difference in mean values of classes is small. NDA focuses on training samples near the required decision boundary, but it does not perform well when either the covariance matrices of the classes are not equal. The main idea of NWFE is to assign different weights to every sample to compute the weighted means and to define new nonparametric between-class and within-class scatter matrices to obtain more than $L-1$ features.

4) Principal Component Analysis (PCA): it is one of the most widely exploited unsupervised approaches in feature reduction. The basic idea of PCA is to find the linearly transformed components that provide the maximum amount of variance possible. Usually the first few components account for a large proportion of the total variance of data and are used to reduce the dimensionality of the original data. However, all components are needed to accurately reproduce the correlation coefficients within the original image. PCA is an unsupervised technique and as such does not include label information for the data.

In the experiments, the software MultiSpec (Purdue Research Foundation) has been used for the NWFE technique, while MATLAB has been used for all the others feature reduction approaches.

D. Classification

1) Support Vector Machine: For each experiment, the performance is reported in terms of classification accuracy, kappa coefficient and the computational time required for the convergence of the ICA. The individual ICA algorithms are compared and analyzed using the three performance measures. For classification purposes an SVM classifier was exploited considering a Radial Basis Function (RBF) kernel. The algorithm exploited is the SVM presented in [42]. The regularization parameter C and the kernel parameter γ are estimated by exploiting a grid-search using a 10-fold cross-validation. This means that the training set is first divided into 10 subsets of equal size, and then each subset is tested using the classifier trained on the remaining 9 subsets. In order to identify the best parameters, exponentially growing sequences of C and γ are considered. In particular, $C = \{10^{-2}, 10^{-1}, ..., 10^{4}\}$ and $\gamma = \{2^{-3}, 2^{-2}, ..., 2^{4}\}$. In this work, the LIBSVM [42] library developed for MATLAB, which employs the one-against-one multiclass strategy, was used.

2) Random Forest: In the experiment II, the classification results obtained by using SVM are compared to the ones obtained by using Random Forest (RF) [43] classifier. This experiment is an exploratory one, in order to validated the role of the classifier in the entire process. Having seen that the pattern of the classifiers behaviours follow similar trends, without loss of generality we conducted the entire experimental design using the SVM classifier. On the contrary of SVM, random forest does not require any parameter tuning. In this work, an implementation developed for MATLAB was used [44].

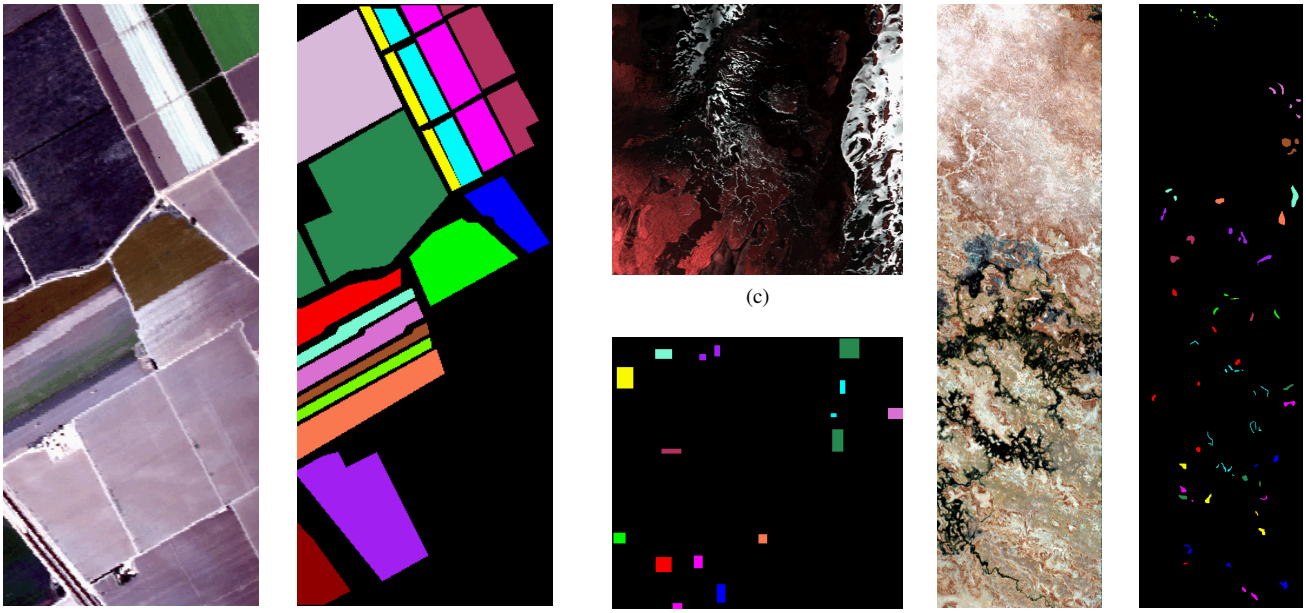
E. Description of the Data Sets

The experimental analysis has been carried out on three real hyperspectral data sets characterized by different spatial and spectral resolutions, and are described below.

1) Salinas Valley, California (Salinas): the data set has been acquired over Salinas Valley, California, in 1998. The acquisition has been done by using the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensor, which uses four spectrometers. The original data set is composed of 224 bands with a spectral range between 0.43 and 0.86 μm. The image has a size of 512 $\times$ 217 pixels with a spatial resolution of 3.7 m. In this study, the corrected data set is considered by discarding the 20 water absorption bands: [108-112], [154-167], 224. The ground reference data contains 16 classes of interest (which are described in Table I). A false color composition of the data set and the reference map are shown in Figs. 1a and 1b. For this data set, the training set used in the experiments is made up of 15% randomly selected samples from each class.

2) Hekla volcano, Iceland (Hekla): the data set was collected in June 17 1991 on the active Hekla volcano, which is located in south-central Iceland, by the 224-band AVIRIS sensor. Due to the failure of the near-infrared spectrometer (spectrometer 4) during the data acquisition, 64 channels appeared blank. After discarding the noisy and the blank channels, the final data set included 157 spectral channels. The image has dimensions of 600 $\times$ 560 pixels with a geometric resolution of 20 m. It shows mainly lava flows from different eruptions and older hyaloclastites (formed during subglacial eruptions). The ground reference data contains 12 classes of interests, which are described in Table I. Figs. 1c and 1d show a false color composition of the image and the reference map, respectively. More information about the data set can be found in [45]. The training set used here was generated by a random selection of 50 samples from each class.

3) Okavango Delta, Botswana (Botswana): the data set was collected over the Okavango Delta, Botswana in 2001-2004 by the Hyperion sensor on EO-1. It acquires data at 30 m resolution over a 7.7 km strip in 242 bands covering the 400-2500 nm portion of the spectrum with a spectral resolution of 10 nm. Uncalibrated and noisy bands that cover water absorption features were removed, and the remaining 145 bands were included as candidate features: [10-55], [82-97], [102-119], [134-164], [187-220]. The considered image is composed of 256 $\times$ 1476 pixels. The ground reference data represent 14 land cover types in seasonal swamps, occasional swamps, and drier woodlands located in the distal portion of the Delta. The data set and the reference map are shown in Figs. 1e and 1f, respectively, while Table I provides information related to the



(a) (b) (c) (d) (e) (f)

Fig. 1. Hyperspectral images (in false color) and the related reference maps. (a), (b) Salinas data set. (c), (d) Hekla data set. (e), (f) Botswana data set.

TABLE I
CLASSES AND NUMBERS OF TRAINING AND TEST SAMPLES FOR EACH DATA SET.

	Salinas Data Set			Hekla Data Set			Botswana Data Set		
No.	Class	Training	Test	Class	Training	Test	Class	Training	Test
1	Broccoli green weeds 1	301	1708	Andesite lava moss cover	50	973	Water	54	216
2	Broccoli green weeds 2	558	3168	Scoria	50	500	Hippo grass	20	81
3	Fallow	296	1680	Hyperclatite formation	50	634	Floodplain grasses1	50	201
4	Fallow rough plow	209	1185	Andesite lava 1980 III	50	1446	Floodplain grasses2	43	172
5	Fallow smooth	401	2277	Rhyolite	50	354	Reeds1	53	216
6	Stubble	593	3366	Andesite lava 1980 I	50	658	Riparian	53	216
7	Celery	536	3043	Andesite lava 1991 II	50	360	Firescar2	51	208
8	Grapes untrained	1690	9581	Andesite lava 1991 I	50	2689	Island interior	40	163
9	Soil vineyard develop	930	5273	Firn and glacier ice	50	408	Acacia woodlands	62	252
10	Corn senesced green weeds	491	2787	Andesite lava 1970	50	292	Acacia shrubland	49	199
11	Lettuce romaine 4 weeks	160	908	Lava with Tephra and Scoria	50	650	Acacia grasslands	61	244
12	Lettuce romaine 5 weeks	289	1638	Snow	50	663	Short mopane	36	145
13	Lettuce romaine 6 weeks	137	779	-	-	-	Mixed mopane	53	215
14	Lettuce romaine 7 weeks	160	910	-	-	-	Exposed soil	19	76
15	Vineyard untrained	1090	6178	-	-	-	-	-	-
16	Vineyard vertical trellis	271	1536	-	-	-	-	-	-

classes. More information about the data set can be found in [46]. The training set used here is generated by random selection of 20% of samples from each class.

For each data set, the training samples and the test samples are generated in such way that the two sets results mutually exclusive (i.e., no shared samples between the two sets).

IV. EXPERIMENTAL RESULTS

In this section we will present and discuss in depth the results of the experiments described above.

A. Experiment I: Low-dimensional space

The results of the analysis conducted in experiment I are depicted in Fig. 2. The taxonomy is based on the different DR approaches, showing for each of them the behaviour of the classification accuracy for the three ICA algorithms considering different subsets of components. Table II reports the best results obtained by the ICA algorithms for each strategy, showing the number of retained components, the overall accuracy (OA), the kappa coefficient (k) and the computational time (CPU time).

For the Salinas data set (Figs. 2, first column), the strategy LFDA-ICA obtains the best classification accuracy. In this case, all the ICA algorithms perform similarly in terms of classification accuracy and number of components required. JADE algorithm outperforms the others in terms of overall accuracy, which reaches a score of 95.48%. However, in terms of computational cost, FastICA requires about 30% less with respect to JADE and 60% less CPU time with respect to Infomax. NWFE-ICA appears to be the second best strategy. Fig. 2g shows that JADE and FastICA provide very

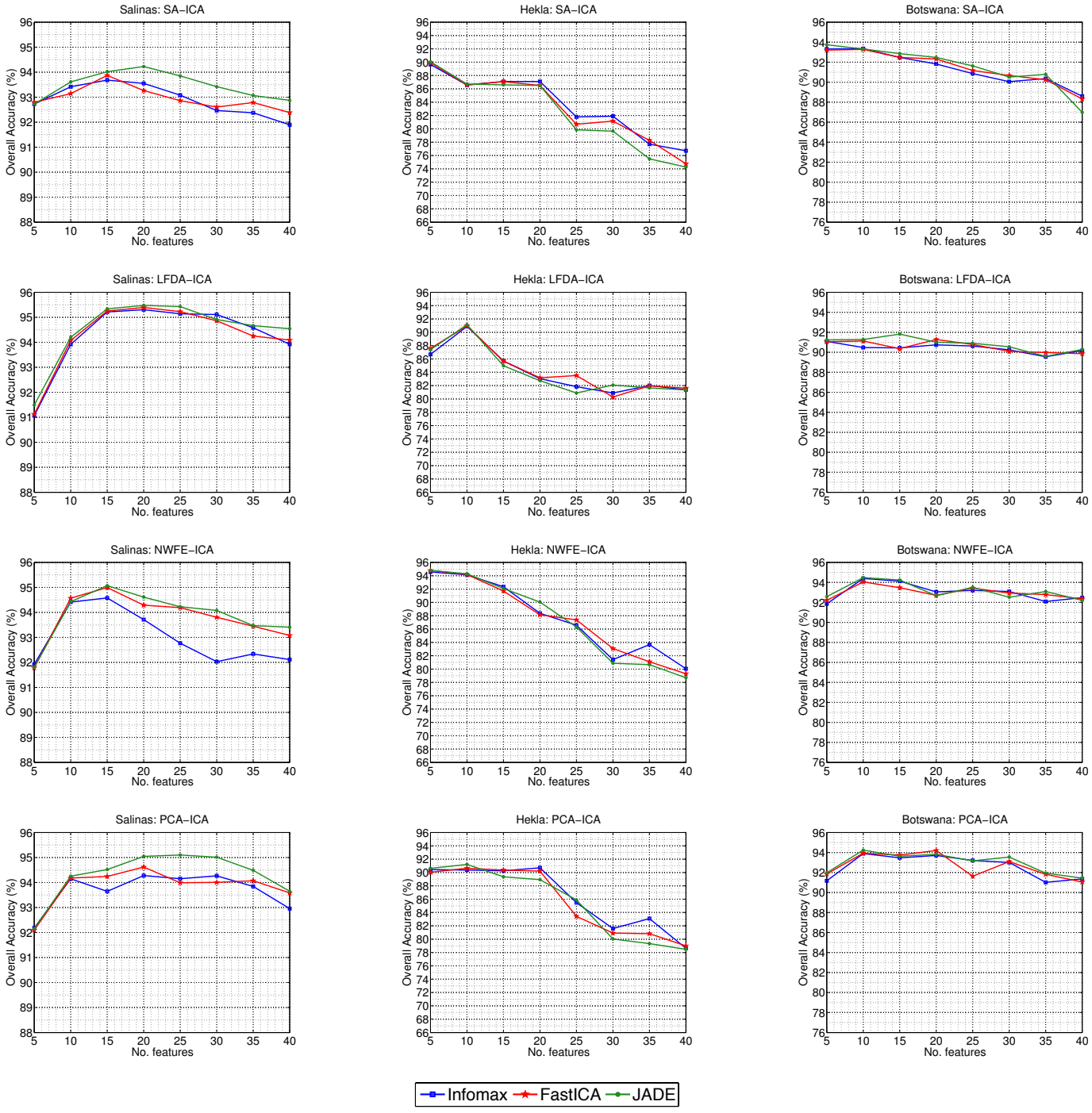


Fig. 2. Experiment I: comparison of the overall classification accuracy obtained by Infomax, FastICA and JADE for different DR strategies (SA-ICA, LFDA-ICA, NWFE-ICA, PCA-ICA) and different number of features: (first column) Salinas data set; (second column) Hekla data set; (third column) Botswana data set.

similar trends, obtaining as highest OA 94.99% and 95.07%, respectively, while the performance of Infomax are strongly affected when increasing the number if ICs. All the ICA algorithms obtain the highest accuracy with 20 components. In terms of computational cost, JADE requires about 20% less CPU time than FastICA and about 75% less than Infomax. In the case of the PCA-ICA strategy, the maximum accuracy obtained by JADE is 95.10% (25 ICs). This requires an higher CPU time than FastICA and Infomax, which provided as highest OA 94.28% and 94.62%, respectively (20 ICs). However, considering the global trend shown in Fig. 2j JADE provides the best classification accuracies with respect to FastICA and Infomax. Also in the case of SA-ICA, JADE provides in general a better global trend with respect to the other two ICAs. In terms of computational time, JADE and FastICA require equal CPU time, while Infomax results to be

TABLE II
CLASSIFICATION RESULTS OBTAINED IN EXPERIMENT I (FIG. 2). ONLY THE BEST RESULTS ARE REPORTED. CLASSIFICATION RESULTS OBTAINED ON THE ORIGINAL SPECTRAL CHANNELS ARE GIVEN FOR COMPARISON. "NO. FEAT." DENOTES THE NUMBER OF FEATURE RETAINED, "OA (%)" DENOTES PERCENTAGE OVERALL ACCURACY, "K" INDICATES THE KAPPA COEFFICIENT AND "TIME" GIVES THE COMPUTATIONAL TIME IN SECONDS.

			SA-ICA			LFDA-ICA			NWFE-ICA			PCA-ICA		
		Spectr.	Infomax	FastICA	JADE	Infomax	FastICA	JADE	Infomax	FastICA	JADE	Infomax	FastICA	JADE
Salinas	No. feat.	204	15	15	20	20	20	20	15	15	15	20	20	25
	OA (%)	94.55	93.68	93.87	94.23	95.30	95.39	**95.48**	94.58	94.99	95.07	94.28	94.62	95.10
	k	0.91	0.93	0.93	0.94	0.90	0.95	0.95	0.94	0.94	0.95	0.93	0.93	0.95
	Time (s)		12.70	7.20	7.22	17.17	7.10	10.56	14.33	4.52	3.57	17.54	7.91	29.70
Hekla	No. feat.	157	5	5	5	10	10	10	5	5	5	10	5	10
	OA (%)	93.89	89.65	89.93	90.06	90.88	90.96	91.14	94.57	94.79	**94.81**	90.64	90.43	91.21
	k	0.80	0.88	0.88	0.86	0.90	0.90	0.90	0.94	0.94	0.94	0.89	0.89	0.90
	Time (s)		44.71	2.38	1.62	34.90	3.40	3.95	26.91	1.42	0.58	32.96	1.87	3.85
Botswana	No. feat.	145	10	10	5	15	20	15	10	10	10	10	10	10
	OA (%)	93.42	93.32	93.28	93.74	90.44	91.28	91.82	94.43	94.00	**94.47**	93.89	93.93	94.24
	k	0.93	0.93	0.93	0.93	0.90	0.91	0.91	0.94	0.94	0.94	0.93	0.94	0.94
	Time (s)		26.90	2.50	0.49	47.50	35.08	18.02	42.28	9.51	3.01	39.58	4.65	4.45

the slowest. Considering the results obtained for Hekla data set (see Figs. 2, second column), the NWFE-ICA approach provides the best results achieving classification accuracies 4% higher than all the other strategies. All the three ICAs provide the best performance when 5 components are retained. The obtained OAs are very close to each other, however, in terms of computational cost, JADE requires about 60% less than FastICA and more than 95% less than Infomax. The LFDA-ICA strategy provides the best OA when 10 components are considered, with an accuracy which is slightly higher than the one obtained with the PCA-ICA. On the other hand, considering different subsets of components, PCA-ICA shows a better trend than the LFDA-ICA. In terms of computational time, JADE and FastICA require similar computational time, which is about 80% less than Infomax. The SA-ICA strategy provided similar (and in some cases better) results with respect to the LFDA-ICA even thought it is the approach that gave the lowest maximum classification accuracy. Each ICA algorithm reached its best performance with 5 components. JADE and FastICA obtained very close results in terms of both OA and CPU time, while Infomax provided a slightly lower OA requiring a much higher computational time. In the analysis of the Botswana data set (Figs. 2, third column), the best accuracies are obtained when applying NWFE-ICA and PCA-ICA. The former provided the highest accuracy when 10 components were considered, where JADE and Infomax provided a slightly higher accuracy than FastICA. However, the computational time required by JADE is about 50% less than FastICA and 90% less than Infomax. A similar analysis can be done for both PCA-ICA and SA-ICA strategies. LFDA-ICA approach provided the lowest classification accuracy with respect to the other strategies. Also in this case, the three ICA algorithms obtain very similar classification accuracies, while the computational time required by JADE is about 50% and 60% smaller than that required by the FastICA and the Infomax, respectively. In general the use of feature extraction algorithms for pre-processing achieves higher classification accuracies than to using feature selection. The reason might be that a better minimization of the noise contribution is achieved when the feature extraction algorithms are employed. Considering the best results reported in the Table II (highlighted in gray), for each data set JADE achieved accuracies that are slightly higher than those of the other ICA algorithms. The highest improvement was achieved in case of Botswana, where JADE outperformed Infomax improving the OA by 1.38%. In terms of computational time required to achieve the best classification accuracy, the JADE's performance is comparable to the one obtained by FastICA. Infomax resulted in general the worst performing technique, both from the computational time and the classification accuracy points of view.

In this experiment, another technique that was recently proposed in the neuroscience field, RobustICA [34], has been used. In [34], RobustICA is presented and compared to the kurtosis-based FastICA, considering the deflationary orthogonalization (i.e., the components are extracted one by one). The study is applied to the biomedical problem of atrial activity (AA) extraction in atrial fibrillation (AF) electrocardiograms (ECGs). In that context, RobustICA was claimed to be more efficient than FastICA in providing better sources with a lower computational cost. In [47] RobustICA was compared to JADE for ECG artifacts removal from Electroencephalogram (EEG) signals. Also in this case, RobustICA was preferred than JADE. Even if the scope of the study is not to exploit all the existing implementations but only the most widely used in the remote sensing field, an exploratory experiment was carried out using the aforementioned implementation. Taking into account these results, the algorithm is tested and compared to the best cases (i.e., LFDA-ICA for Salinas, NWFE-ICA for Hekla and Botswana). Fig. 3 shows the comparison between the ICA algorithms, while the best obtained results are reported in Table III From the experimental analysis, it can be seen that in terms of classification accuracies, the performance of RobustICA is in line with the ones obtained by FastICA, Infomax and JADE approaches, providing the best accuracy among the other ICA algorithms in case of Botswana. However, the required computational cost is much higher, resulting in a extremely slow computational time (especially in case of Botswana), which seems to increase linearly with

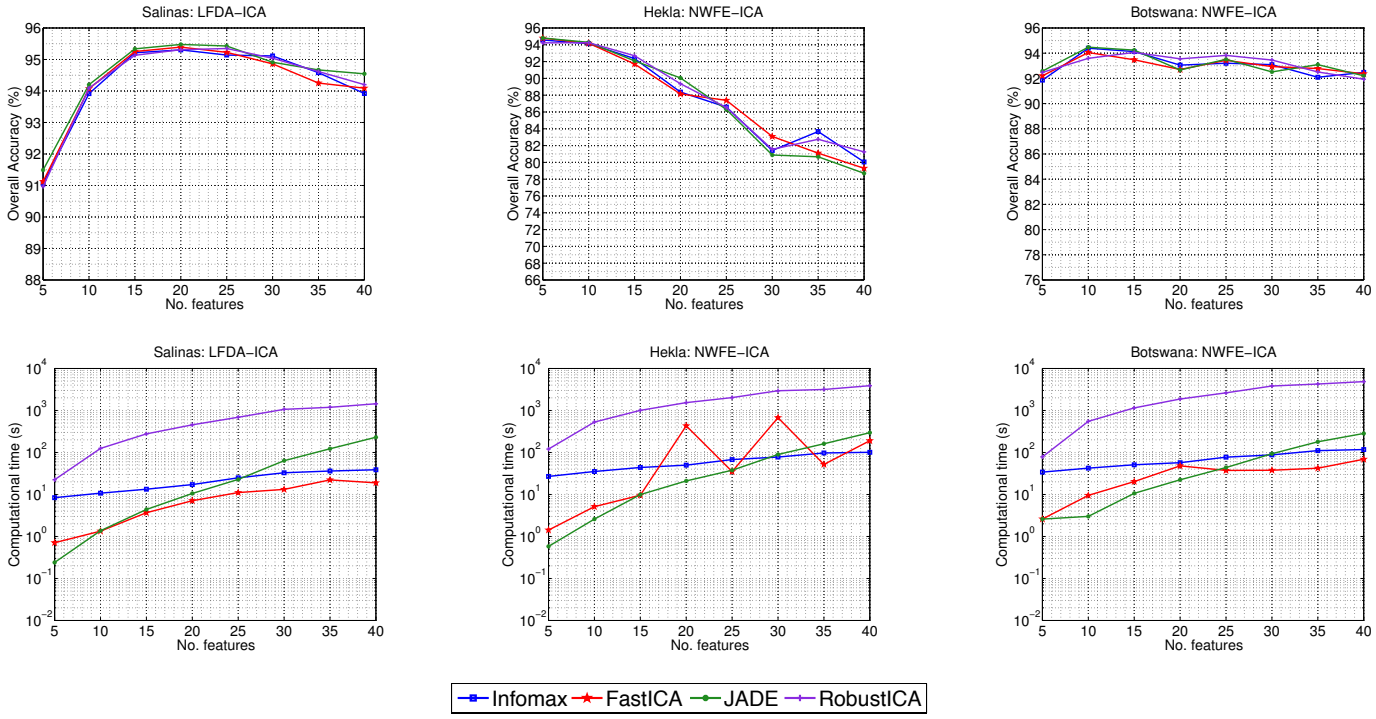


Fig. 3. Experiment I: comparison of the overall classification accuracy and computational cost obtained by Infomax, FastICA, JADE and RobustICA versus the number of features, considering the best DR strategies: (first column) LFDA-ICA for Salinas data set, (second column) NWEFE-ICA for Hekla data set; (third column) NWFE-ICA for Botswana data set.

TABLE III
CLASSIFICATION RESULTS OBTAINED IN EXPERIMENT I CONSIDERING ROBUSTICA ALGORITHM AND THE BEST DR STRATEGIES (FIG. 3). ONLY THE BEST RESULTS ARE REPORTED. "NO. FEAT." INDICATES THE NUMBER OF FEATURE RETAINED, "OA (%)" DENOTES PERCENTAGE OVERALL ACCURACY, "K" GIVES THE KAPPA COEFFICIENT AND "TIME" GIVES THE COMPUTATIONAL TIME IN SECONDS.

	Salinas: LFDA-ICA				Hekla: NWFE-ICA				Botswana: NWFE-ICA			
	Infomax	FastICA	JADE	RobustICA	Infomax	FastICA	JADE	RobustICA	Infomax	FastICA	JADE	RobustICA
No. feat.	20	20	20	25	5	5	5	10	10	10	10	15
OA (%)	95.30	95.39	95.48	95.35	94.57	94.79	94.81	94.25	94.43	94.00	94.47	94.59
k	0.90	0.95	0.95	0.95	0.94	0.94	94	0.94	0.94	0.94	0.94	0.94
Time (s)	17.17	7.10	10.56	688.74	26.91	1.42	0.58	529.01	42.28	9.51	3.01	1151.5

the number of extracted components. This methods have been used in neuroscience field but never in remote sensing field. However, the nature and the properties of the hyperspectral images are different from signals analyzed in neuroscience field. Considering the obtained results and because of the significant computational cost required even when feature reduction is performed in pre-processing, the use of these techniques does not seem appropriate for hyperspectral images. Thus it will not be considered further in the experimental analysis.

B. Experiment II: High-dimensional space

Following the design of the experiment I, also in this case the overall accuracies are entirely depicted in Fig. 4, while the best results are reported in Table IV. This Table shows for each chosen algorithm (in this case only Infomax and FastICA), the number of retained components, the overall accuracy (OA), the kappa coefficient (k) and the computational time, which is the estimation cost of the ICA on the entire original data set. Both of the results obtained by using SVM and RF classifiers are reported. From the analysis of the results obtained by using SVM, it can be noticed that FastICA outperforms Infomax for all the three data sets, increasing the OA by 2.16% in the case of Salinas and by 1.33% in the case of Botswana. In the case of Hekla the results are quite similar, showing a little improvement of 0.56% for FastICA. The kappa coefficient follows a similar trend as well. FastICA requires a CPU timewhich is one order of magnitude less with respect to Infomax, confirming the superiority of the fixed-point algorithm. The use of the entire data set, without performing dimensionality reduction, assures that there is no information loss in the process, as it may happen when any feature extraction/selection technique is used. Considering the case of Salinas, the selection of 35 ICs provided an OA of 94.12%, which is quite close to the one obtained in the first experiment by the SA-ICA approach (94.23%) (see in Table II). However by applying ICA to the entire data set there is no noise reduction, the extracted components carry noise which affects the final classification results. This becomes

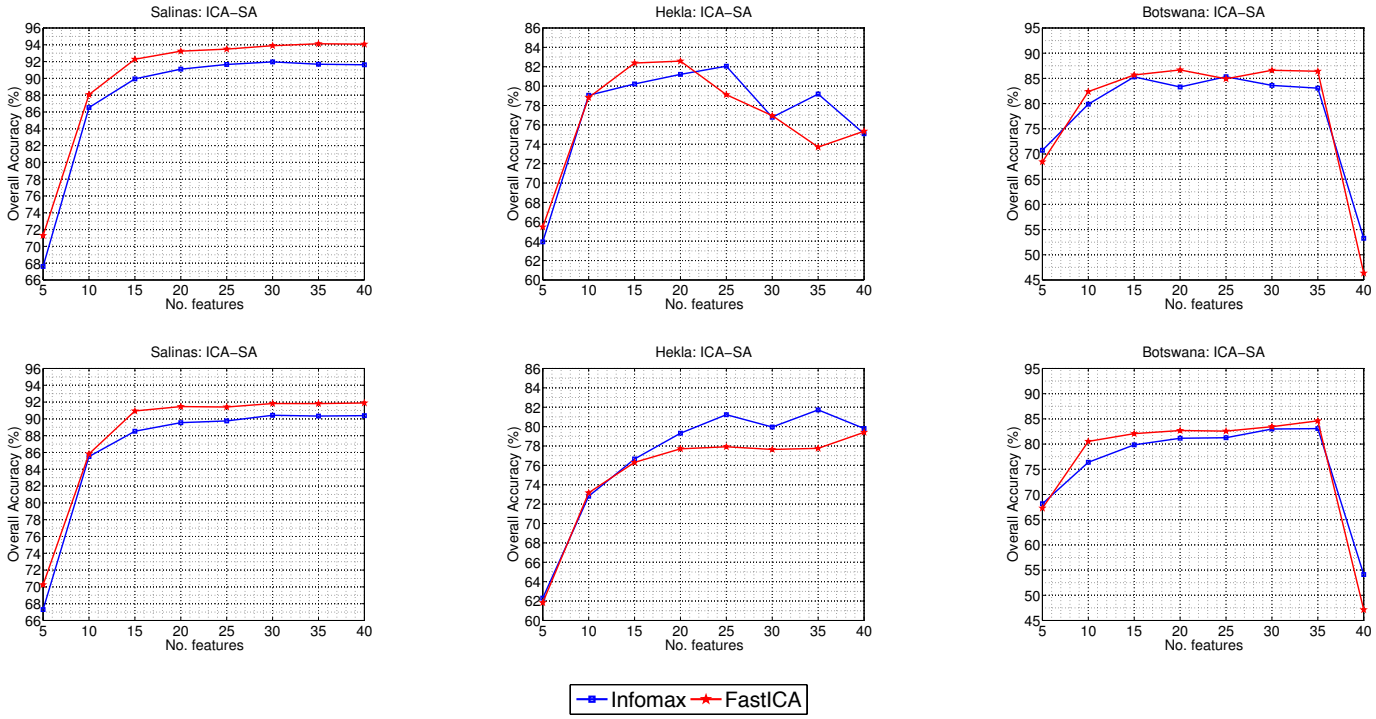


Fig. 4. Experiment II: comparison of the overall classification accuracy obtained by Infomax and FastICA versus the number of features by using (top row) SVM, and (bottom row) RF, for Salinas, Hekla and Botswana data sets.

TABLE IV
CLASSIFICATION RESULTS OBTAINED IN EXPERIMENT II. ONLY THE BEST RESULTS ARE REPORTED. "NO. FEAT." INDICATES THE NUMBER OF FEATURE RETAINED, "OA (%)" DENOTES PERCENTAGE OVERALL ACCURACY, "K" GIVES THE KAPPA COEFFICIENT AND "TIME" GIVES THE COMPUTATIONAL TIME IN SECONDS.

		Salinas		Hekla		Botswana	
		Infomax	FastICA	Infomax	FastICA	Infomax	FastICA
SVM	No. feat.	30	35	25	20	25	20
	OA (%)	91.96	94.12	82.05	82.58	85.33	86.67
	k	0.91	0.93	0.80	0.80	0.84	0.86
RF	No. feat.	30	40	35	40	35	35
	OA (%)	90.43	91.89	81.73	79.41	83.06	84.60
	k	0.89	0.91	0.79	0.77	0.82	0.83
	Time (s)	2253.16	281.22	3694.57	2506.31	3028.29	337.19

more evident in the experimental results obtained for the Hekla and Botswana data sets, in which the accuracies are quite low with respect to the ones obtained in the first experiment.

Considering the results obtained by using RF, the overall accuracies are in general lower than the ones of SVM. However, the obtained results follow a similar trend of the ones obtained with SVM in case of Salinas and Botswana, where FastICA outperform Infomax, while in case of Hekla, Infomax is the one that obtained the highest overall accuracy.

C. Experiment III: Spatial downsampling

In this experiment the performances of the ICA algorithms are investigated and compared when a spatial downsampling of the ICA's imput data is performed. The analysis presents the results obtained by considering two scenarios.

1) Low-dimensional space: In this experiment, the three ICA algorithms were tested on the three different data sets taking into account only the *DR-approach-ICA* strategies that gave the best results in terms of accuracies in the experiment I (see Table II), i.e., the LFDA-ICA in case of Salinas data set, and the NWFE-ICA in the case of both Hekla and Botswana data sets. Table V reports the number of samples employed in the experiment. *None (all samples)* denotes the case in which the entire image is considered. This coincides with the results obtained in experiment I, and they are reported here for comparison. *Ds2x*, *Ds3x*, *Ds4x* denote a decrease of the sampling rate of a factor 2, 3 and 4, respectively. *Train 1* denotes the initial training sets, i.e., 15% of the ground truth samples in the case of Salinas, 50 samples for each class in case of Hekla, and 20% of the ground truth in case of Botswana. *Train 2* denotes 10% of the ground truth in

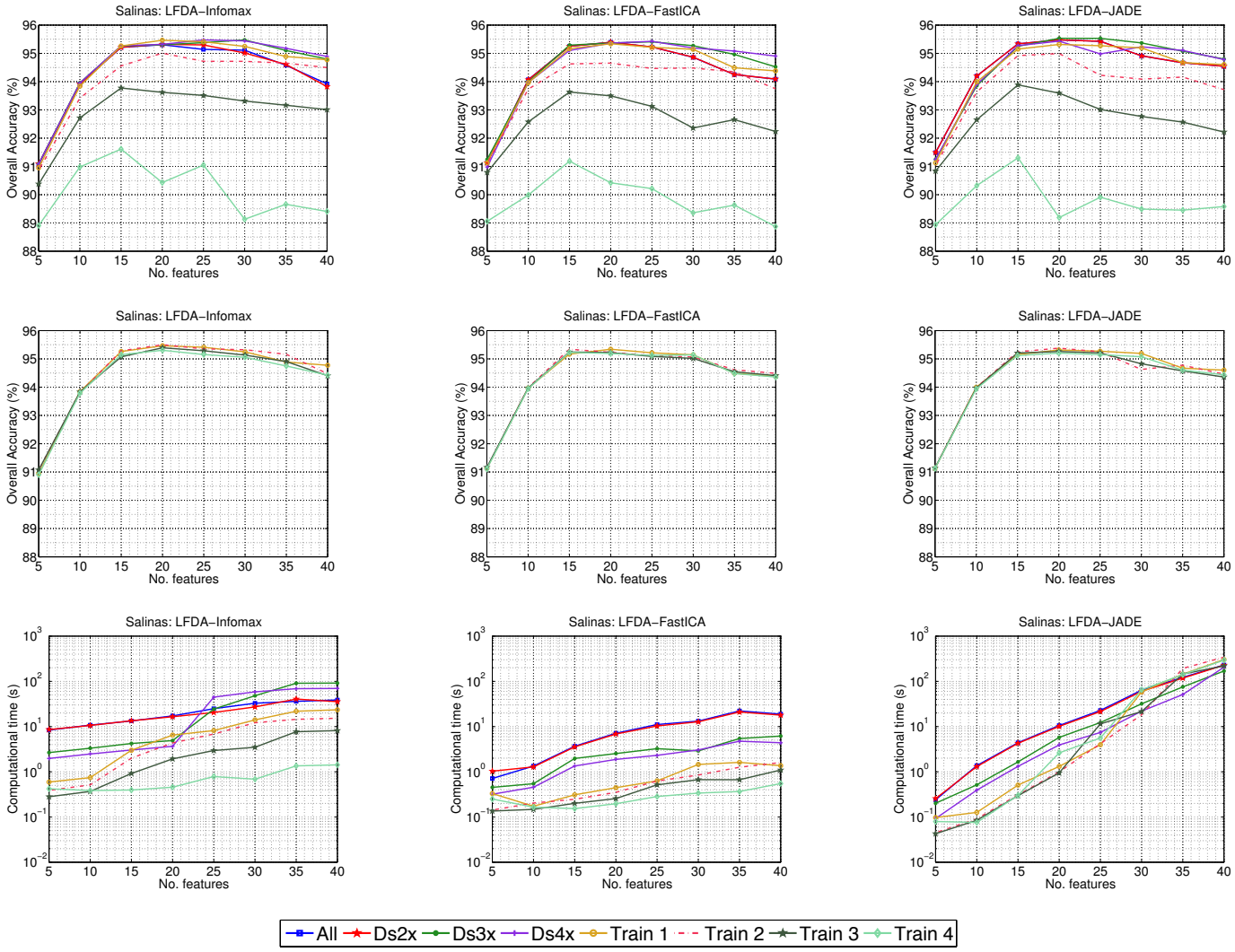


Fig. 5. Experiment III in low dimensional scenario: comparison of the overall classification accuracy provided by (first column) Infomax, (second column) FastICA, (third column) JADE, for different number of samples on Salinas data set. Top row shows the results obtained by exploiting the first approach (i.e, the same number of training samples are given as input to both the ICA and the classifier), while the middle row shows the ones obtained by exploiting the second approach (i.e, the number of the training samples given as input to the ICA varies, while the one in input to the classifier remains the same). The bottom row shows the computational time related to the first approach.

case of Salinas, 40 samples for each class in the case of Hekla, and 15% of the ground truth samples in the case of Botswana. *Train 3* denotes 5% of the ground truth in case of Salinas, 30 samples for each class in the case of Hekla, and 10% of the ground truth samples in the case of Botswana. *Train 4* denotes 2% of the ground truth in case of Salinas, 10 samples for each class in the case of Hekla, and 5% of the ground truth samples in the case of Botswana. For an easier interpretation of the global behaviours of the ICA algorithms, the results are reported as graphs in Figs. 5-7, which show the OAs and the computational time obtained when different downsampling factors and training samples are considered. For a more exhaustive analysis of the effect of the reduction of the training samples on the performance of the ICA algorithms, two different approaches are considered. The first approach takes into account a real life situation, where the same number of training samples for the classifier and the ICA is considered, In this case, the observed variation of the classification accuracy trend is caused of the degradation of both the ICs and the effectiveness of the classifier. In the second approach, the number of the training samples in input to the ICA varies, while the one in input at the classifier (which

TABLE V
EXPERIMENT III: DESCRIPTION OF THE DATA SET CONSIDERED IN TERMS OF NUMBERS OF THE SAMPLES.

Downsampling	Salinas	Hekla	Botswana
None (all samples)	111104	336000	377856
Ds2x	55552	168000	188928
Ds3x	37035	112000	125952
Ds4x	27776	84000	94464
Train 1	8112	600	644
Train 2	5403	480	482
Train 3	2697	360	323
Train 4	1075	120	161

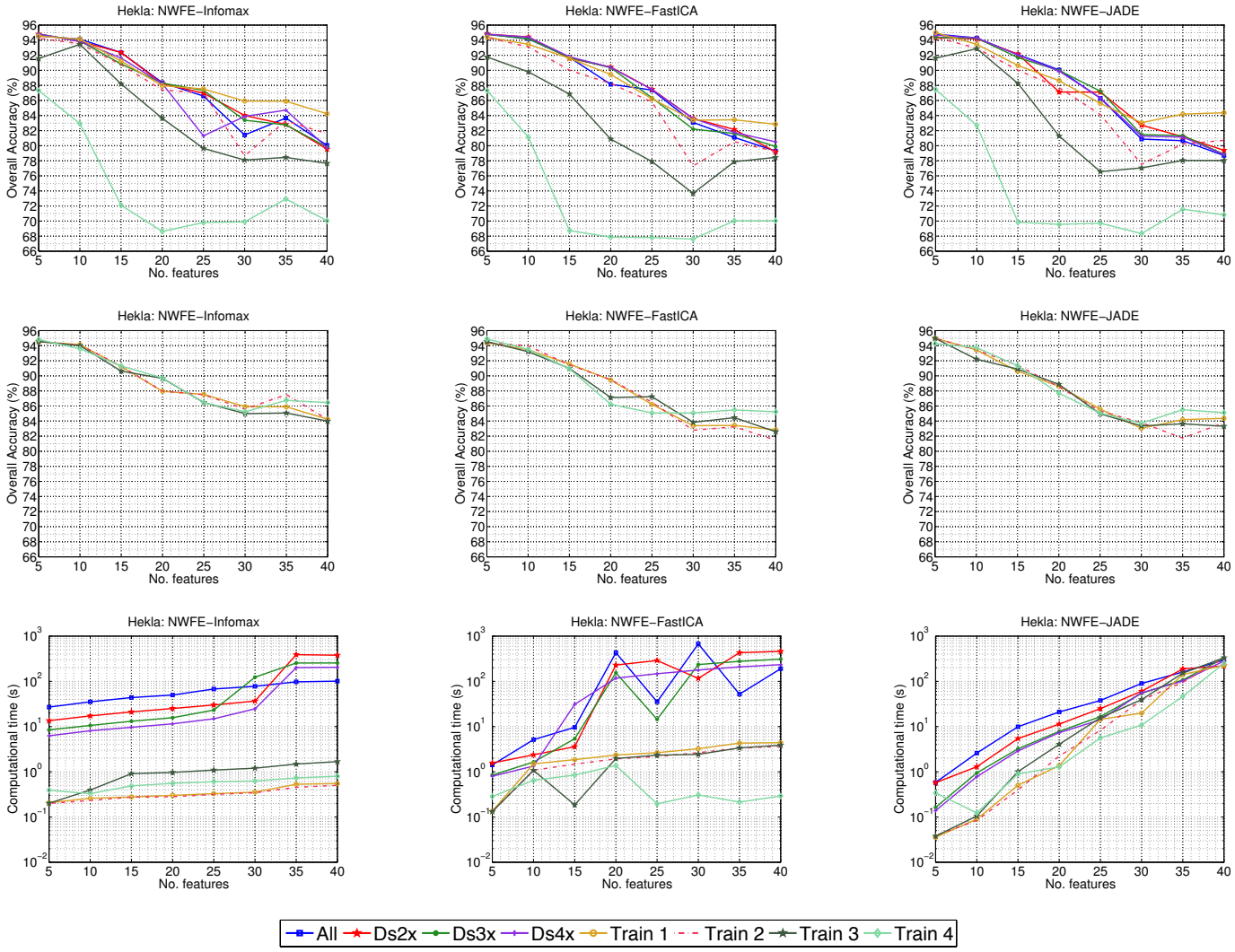


Fig. 6. Experiment III in low dimensional scenario: comparison of the overall classification accuracy provided by (first column) Infomax, (second column) FastICA, (third column) JADE, for different number of samples on Hekla data set. Top row shows the results obtained by exploiting the first approach (i.e, the same number of training samples are given as input to both the ICA and the classifier), while the middle row shows the ones obtained by exploiting the second approach (i.e, the number of the training samples given as input to the ICA varies, while the one in input to the classifier remains the same). The bottom row shows the computational time related to the first approach.

coincides with the original training set (i.e., *Train. 1*)), remains the same. Even if this strategy is unusual, it permits us to evaluate the real effect of the reduction of the training samples on the performance of the ICA. For each of the Figs. 5, 6 and 7, the top row reports the results obtained considering the first approach, the middle row reports the results obtained by the second approach and the bottom row shows the log-lin plots of the computational times (plotted in logarithmic scale on the y-axis) for each of the sub-experiment reported in the top row. The analysis is done for different numbers of components (plotted in linear scale on the x-axis) retrieved each time.

Focusing on the Salinas data set, the performances of each ICA technique, obtained by applying a downsampling of a factor 2, are very close to the ones obtained by using all the samples. Similarly, the computational times required for the convergence are similar. Improvements in the overall accuracy, and especially in the computational time, are more evident when higher factors are considered. In particular, when Infomax is used, the best OA (95.48%) is obtained when the sampling rate is decreased by a factor 3 and 4. While the OAs are quite similar, the computational times improve with respect to the case in which all the samples are used. For example the computational time decreased to 6.47 *s* by using only the training samples, achieving an OA of 95.46%. Considering FastICA, the best OA (95.43%) was achieved by applying a downsampling of a factor 4, while the obtained computational time was 2.31 *s*. For JADE, the highest OA (95.53%) is obtained by reducing the sample rate by a factor 3, halving the computational time (5.74 *s*) with respect to the case when the entire data set is used (10.56 *s*).

Analyzing the results obtained for the Hekla data set, the downsampling reduces the computational times. For Infomax, 6.25 *s* is the computational time reached with a downsampling of a factor 4, which is a much lower compared to the 26.91

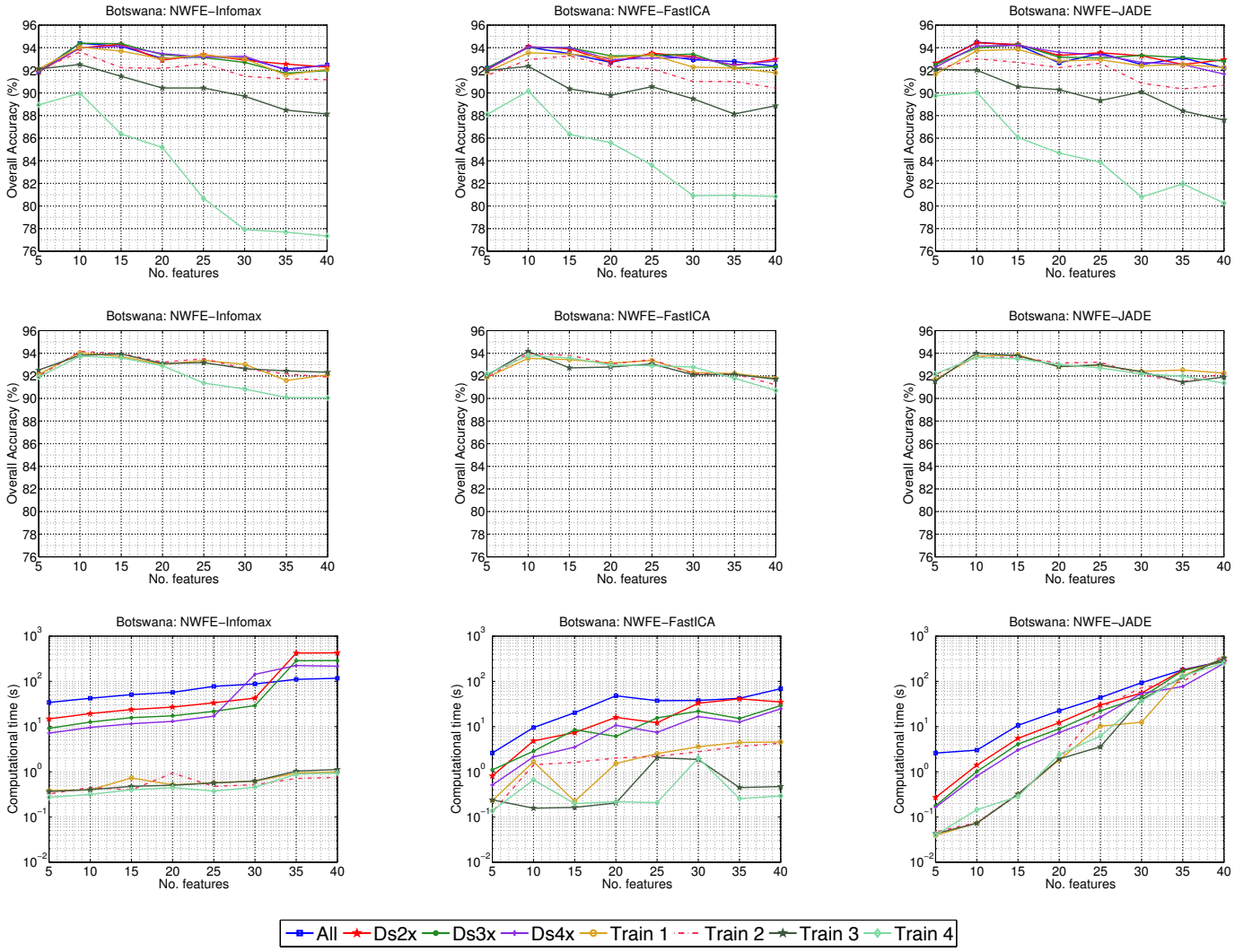


Fig. 7. Experiment III in low dimensional scenario: comparison of the overall classification accuracy provided by (first column) Infomax, (second column) FastICA, (third column) JADE, for different number of samples on Botswana data set. Top row shows the results obtained by exploiting the first approach (i.e, the same number of training samples are given as input to both the ICA and the classifier), while the middle row shows the ones obtained by exploiting the second approach (i.e, the number of the training samples given as input to the ICA varies, while the one in input to the classifier remains the same). The bottom row shows the computational time related to the first approach.

s obtained in the experiment I. While time improved, the OA resulted 94.83%. The classification performances obtained by using the FastICA remain the same for all the sub-experiments, while the computational time decreased from 1.42 *s* (obtained in the experiment I) to 0.80 *s*. The JADE algorithm improved the computational time, which decreased from 0.58 *s* to 0.036 *s* when using only the training samples, achieving an OA of 95.00%, which is slightly higher than the previous OA of 94.81%.

For Botswana data set, the downsampling does not improve the performances of JADE, which achieved a slightly lower OA (94.28 % with 15 components) than the one obtained in the experiment I (94.47% with 10 components), with a computational time that slightly increased to 4.1 *s*. A different behaviour can be noticed for Infomax and FastICA. Both of them decreased the computational cost, which was reduced by 77% (from 42.28 *s* to 9.57 *s*) for Infomax, and by 50% (from 9.51 *s* to 4.81 *s*) for FastICA, without decreasing the classification accuracies (from 94.43% to 94.51% for Infomax, and from 94.00% to 94.05% for FastICA).

Comparing the only variation in the number of training samples, it can be seen that a decrease of the number of training samples in the first approach (top row) strongly affects the classification accuracy. The analysis performed by varying only the number of training samples given as input to the ICA (middle row), keeping constant the number of the training samples in input to the classifier, shows that the quality of the extracted ICs is not affected, providing similar trend of classification accuracy for a different number of training samples. This result points out that a variation in the number of the training samples affects more incisively the classifier rather than the ICA performance. In terms of computational time, decreasing the number of training samples coincides in general with a decrease of the computational cost. This is more

TABLE VI
CLASSIFICATION RESULTS OBTAINED IN EXPERIMENT III CONSIDERING THE HIGH DIMENSIONAL SCENARIO. THE RESULTS ARE RELATED TO SALINAS DATA SET BY USING FASTICA. "NO. FEAT." INDICATES THE NUMBER OF FEATURE RETAINED, "OA (%)" DENOTES PERCENTAGE OVERALL ACCURACY, "K" GIVES THE KAPPA COEFFICIENT AND "TIME" GIVES THE COMPUTATIONAL TIME IN SECONDS.

	Downsampling x 2 (55552 samples)		Downsampling x 3 (37035 samples)		Downsampling x 4 (27776 samples)		Only training samples (8112 samples)	
No. feat.	OA (%)	k	OA (%)	k	OA (%)	k	OA (%)	k
5	69.25	0.65	65.98	0.62	65.36	0.61	62.48	0.57
10	87.53	0.86	85.32	0.84	85.22	0.84	84.13	0.82
15	92.43	0.92	89.57	0.88	90.39	0.89	90.45	0.89
20	93.55	0.93	90.91	0.90	91.35	0.90	91.57	0.91
25	93.72	0.93	92.17	0.91	92.06	0.91	**91.59**	0.91
30	93.78	0.93	92.01	0.91	**92.63**	0.91	91.27	0.90
35	**94.21**	0.94	**92.62**	0.92	91.34	0.90	90.57	0.90
40	93.97	0.93	91.17	0.90	90.89	0.90	90.15	0.90
Time (s)	102.84		84.57		208.88		161.46	

evident in the case of Infomax and FastICA, while in case of JADE the computational cost increases with the number of retained components.

In general, it can be stated that when dimensionality reduction is performed, downsampling as a pre-processing approach can contribute to the improvement of the computational time of the ICA algorithms without decreasing the overall classification accuracy. This finding is significant, especially when the computational time is an important aspect of the analysis, as is the case of the analysis of hyperspectral images.

2) High-dimensional space: The results reported in Table VI show the performances of ICA when a spatial downsampling of the image is performed before applying ICA in the high-dimensional space scenario are considered. The convergence capability of the ICA is strongly affected by the decreased number of input samples when applied to the entire image, while it fails to converge when very few input samples are considered. For this reason we consider only the results obtained by using the first five data sets described in Table V. For similar reasons, only the results obtained by applying FastICA on Salinas data set are reported. Table VI reports the OA and the k coefficient for different subsets of features, while the computational time is referred to the total time requested for the extraction of all the ICs (i.e., 204 ICs). It can be seen that performing a downsampling of factor a 2 (meaning that half of the total number of samples are discarded), the obtained overall classification accuracy (94.21%) is quite similar to the one achieved by considering all the samples (94.12%, see Table IV), whereas the computational time for extracting the ICs decreases by $63,4\%$. However, when the downsampling factor increases, the performance of the ICA decreases, i.e., the classification accuracy decreases while the required computational time to converge increases.

V. DISCUSSION AND CONCLUSION

In this paper a detailed comparison among three widely used ICA algorithms (i.e., Infomax, FastICA and JADE) for hyperspectral image classification was presented. The analysis took into account different scenarios in order to compare and identify the best strategy for extracting class-discriminant components based on the use of ICA. The ICA algorithms were tested in both low and high dimensional spaces.

In the first scenario, ICA algorithms were tested performing dimensionality reduction with alternative strategies rather than the PCA (which usually is implemented in conjunction with the ICA algorithms and used for retaining a certain number of components). Supervised feature selection/extraction techniques have been exploited and compared to the case in which PCA is used. The results of the analysis point out that the exploitation of prior information in feature extraction methods for dimensionality reduction allows ICA algorithms to provide better feature sets which led to more accurate classifications. In this scenario, which is the most common in the analysis of hyperspectral images, JADE was the ICA approach that provided the best performance in terms of classification accuracy, while it provided results comparable to the ones obtained by FastICA in terms of computational time (in many cases it was faster). Infomax resulted in general to be the worst in terms of both computational time and classification accuracy.

The second scenario was aimed at investigating the performance of ICA when the entire data set is considered. Using the entire data set, without applying any dimensionality reduction, assures that no information is lost before performing the ICA. The analysis in this scenario showed that FastICA outperforms Infomax both in terms of computational time and classification accuracy. In this case, JADE could not be exploited since it requires a massive computational load when the number of estimated components becomes high. When the entire data set is considered, there is theoretically no information loss. However, the full set of selected components is more noisy, thus affecting the classification results.

The third scenario showed that the reduction of the number of samples on which applying ICA can in general improve the ICA convergence speed, without decreasing the classification accuracies. The approach is more effective in low dimensional spaces, where there are no issues with the Hughes' phenomenon, especially when the number of training samples given as input to both the ICA algorithm and the classifier are chosen properly. Indeed, the experiments showed that SVM is more affected by a decrease of the number of the training samples than the ICA, which can provide "good" ICs even when few samples are exploited for the transformation.

This observation becomes very important in applications for which the computational time and the number of available samples are crucial aspects. Consequently, the inclusion of the analysis of prior information in computational efficient strategies should foster the development of new ICA-based methodologies for the analysis of large hyperspectral remote sensing images.

APPENDIX

ACCURACY ASSESSMENT

From the analysis of the confusion matrix of the classification result it is possible to derive useful parameters that indicate how good the obtained classification is. The parameters that are used in this work are the overall accuracy and the kappa coefficient.

Overall Accuracy

The overall accuracy (OA) represents the number (or percentage) of pixels that are correctly classified. Considering a total of C classes, the OA is mathematically defined as the ratio between the total number of corrected pixels for each class, N_i (which are summed along the major diagonal), divided by the total number of referenced pixels that are being tested, T:

$$OA = \frac{\sum_i^C N_i}{T}. \tag{13}$$

Kappa Coefficient

The kappa coefficient (or kappa statistic) provides a measure of overall classification quality by comparing the agreement against the one expected by chance. It is mathematically defined as follows:

$$k = \frac{m_o - m_c}{1 - m_c}, \tag{14}$$

where m_o represents the proportion of correct agreement in the test set, and m_c is the proportion of agreement that is expected by chance. The possible values range from +1 (perfect agreement) via 0 (no agreement above that expected by chance) to -1 (complete disagreement).

ACKNOWLEDGMENTS

This research was in part supported by the research funds of the University of Iceland and in part by the research funds of the University of Trento.

REFERENCES

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

[2] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., ser. Computer science and scientific computing, W. Rheinboldt, Ed. Academic Press, Inc., 1990, vol. 2, no. 1.

[3] S. Lim, K. Sohn, and C. Lee, "Principal component analysis for compression of hyperspectral images," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS '01)*, vol. 00, no. C, 2001, pp. 97–99.

[4] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, 1988.

[5] C. Lee and D. A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 388–400, 1993.

[6] L. O. Jimenez and D. A. Landgrebe, "Hyperspectral data analysis and supervised feature reduction via projection pursuit," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2653–2667, 1999.

[7] B. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 5, pp. 1096–1105, 2004.

[8] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, Inc., 2001.

[9] S. Makeig, A. J. Bell, T. P. Jung, and T. J. Sejnowski, "Independent Component Analysis of Electroencephalographic Data," *Advances in Neural Information Processing Systems*, pp. 145–151, 1996.

[10] T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–78, Mar. 2000.

[11] R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings." *IEEE Transactions on Bio-Medical Engineering*, vol. 47, no. 5, pp. 589–93, May 2000.

[12] S. D. Parmar, H. K. Patel, and J. S. Sahambi, "Separation performance of ICA algorithms on FECG and MECG signals contaminated by noise," *Applied Computing*, vol. 3285, pp. 184–190, 2004.

[13] A. Subasi and M. Ismail Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8659–8666, Dec. 2010.

[14] A. Cichocki and S.-i. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, Inc., 2002.

[15] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1586–1600, 2006.

[16] J. M. P. Nascimento and J. M. B. Dias, "Does independent component analysis play a role in unmixing hyperspectral data?" *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 175–187, Jan. 2005.

[17] J. Wang and C.-I. Chang, "Applications of Independent Component Analysis in Endmember Extraction and Abundance Quantification for Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2601–2616, 2006.

[18] J. A. Palmason, J. A. Benediktsson, J. R. Sveinsson, and J. Chanussot, "Classification of hyperspectral data from urban areas using morpholgical preprocessing and independent component analysis," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS '05)*, vol. 1, pp. 176–179, 2005.

[19] C. Jutten, S. Moussaoui, and F. Schmidt, "How to Apply ICA on Actual Data ? Example of Mars Hyperspectral Image Analysis," in *15th International Conference on Digital Signal Processing*. IEEE, Jul. 2007, pp. 3–12.

[20] Q. Du, I. Kopriva, and H. Szu, "Independent-component analysis for hyperspectral remote sensing imagery classification," *Optical Engineering*, vol. 45, no. 1, Jan. 2006.

[21] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral Image Classification With Independent Component Discriminant Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 4865–4876, 2011.

[22] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of Hyperspectral Images by Using Extended Morphological Attribute Profiles and Independent Component Analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 3, pp. 542–546, 2011.

[23] J. Xia, P. Du, X. He, and J. Chanussot, "Hyperspectral Remote Sensing Image Classification Based on Rotation Forest," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 1, pp. 239–243, 2014.

[24] N. Correa and T. Adali, "Comparison of blind source separation algorithms for fmri using a new matlab toolbox: GIFT," in *ICASSP*, 2005, pp. 401–404.

[25] K. E. Hild, G. Alleva, S. Nagarajan, and S. Comani, "Performance comparison of six independent components analysis algorithms for

fetal signal extraction from real fMCG data." *Physics in medicine and biology*, vol. 52, no. 2, pp. 449–62, Jan. 2007.
[26] A. Kachenoura, L. Albera, L. Senhadji, and P. Comon, "ICA: a potential tool for BCI systems," *IEEE Signal Processing Megazine*, no. January 2008, pp. 57–68, 2008.
[27] K. Vanderperren, M. De Vos, J. R. Ramautar, N. Novitskiy, M. Mennes, S. Assecondi, B. Vanrumste, P. Stiers, B. R. H. Van den Bergh, J. Wagemans, L. Lagae, S. Sunaert, and S. Van Huffel, "Removal of BCG artifacts from EEG recordings inside the MR scanner: a comparison of methodological and validation-related aspects." *NeuroImage*, vol. 50, no. 3, pp. 920–34, Apr. 2010.
[28] A. J. Bell and T. J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995.
[29] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis." *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–34, Jan. 1999.
[30] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Radar and Signal Processing, IEE Proceedings F*, vol. 140, no. 6, pp. 362–370, 1993.
[31] A. Cheriyadat and L. M. Bruce, "Why principal component analysis is not an appropriate feature extraction method for hyperspectral data," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS '03)*, vol. 6, no. C, 2003, pp. 3420–3422.
[32] C.-I. Chang and Q. Du, "Estimation of Number of Spectrally Distinct Signal Sources in Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608–619, 2004.
[33] S. Prasad and L. M. Bruce, "Limitations of Principal Components Analysis for Hyperspectral Target Recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 625–629, 2008.
[34] V. Zarzoso and P. Comon, "Robust Independent Component Analysis by Iterative Maximization of the Kurtosis Contrast With," *IEEE Transactions on Neural Networks*, vol. 21, no. 2, pp. 248–261, 2010.
[35] S. Makeig, A. J. Bell, T. P. Jung, and T. J. Sejnowski, "EEGLAB: A Matlab toolbox for electrophysiological research," 2002. [Online]. Available: http://sccn.ucsd.edu/eeglab/
[36] S. Amari, A. Cichocki, and H. H. Yang, "A New Learning Algorithm for Blind Signal Separation," *Advances in Neural Information Processing Systems*, pp. 757–763, 1996.
[37] S. B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1360–1367, 2001.
[38] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179–188, 1936.
[39] X. He and P. Niyogi, "Locality preserving projections," *Neural information processing systems*, vol. 16, p. 153, 2004.
[40] M. Sugiyama, "Local Fisher discriminant analysis for supervised dimensionality reduction," in *Proceedings of the 23rd international conference on Machine learning - ICML '06*. New York, New York, USA: ACM Press, 2006, pp. 905–912.
[41] K. Fukunaga and J. Mantock, "Nonparametric discriminant analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, no. 6, pp. 671–678, 1983.
[42] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, Apr. 2011.
[43] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.
[44] A. Jaiantilal, "Classification and Regression by randomForest-matlab," Available at \url{http://code.google.com/p/randomforest-matlab}, 2009.
[45] J. A. Benediktsson, "Statistical and Neural Network Pattern Recognition Methods for Remote Sensing Applications," in *Handbook of pattern recognition and computer vision*, 2nd ed., C. H. Chen, L. F. Pau, and P. S. P. Wang, Eds. Singapore: World Scientific Press, 1999, pp. 507–534.
[46] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *Geoscience and Remote and Remote Sensing*, vol. 43, no. 3, pp. 492–501, 2005.
[47] V. Matic, W. Deburchgraeve, and S. Van Huffel, "Comparison of ICA algorithms for ECG artifact removal from EEG signals," in *Proc. of the 4th Annual symposium of the IEEE-EMBS Benelux Chapter (IEEE-EMBS).*, 2009, pp. 2–5.

Nicola Falco (S'10) received the B.Sc. and M.Sc. degrees in telecommunication engineering from the University of Trento, Trento, Italy, in 2008 and 2011, respectively, and since 2011, he has been pursuing the Ph.D. degree (joint degree) at the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento, and with the Signal Processing Laboratory, Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavík, Iceland. From April to October 2011, he served as a Research Assistant with the Signal Processing Laboratory, Faculty of Electrical and Computer Engineering, University of Iceland. His research interests include remote sensing image processing, mathematical morphology, and pattern recognition, including analysis and classification in hyperspectral imagery, as well as change detection analysis in optical remote sensing data.

Mr. Falco serves as a Reviewer for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and *Pattern Recognition Letters*. He was a recipient of the Recognition of IEEE GEOSCIENCE AND REMOTE SENSING LETTERS Best Reviewers in 2013.

Jón Atli Benediktsson (S'84-M'90-SM'99-F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, in 1984, and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, in 1987 and 1990, respectively. He is currently Pro Rector for Academic Affairs and Professor of Electrical and Computer Engineering at the University of Iceland. His research interests are in remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing, and he has published extensively in those fields. Prof. Benediktsson was the 2011-2012 President of the IEEE Geoscience and and Remote Sensing Society (GRSS) and has been on the GRSS AdCom since 2000. He was Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008 and has served as Associate Editor of TGRS since 1999, the IEEE Geoscience and Remote Sensing Letters since 2003 and IEEE Access since 2013. He is on the editorial board of Proceedings of the IEEE and an International Editorial Board of the *International Journal of Image and Data Fusion* and was the Chairman of the Steering Committee of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (J-STARS) 2007-2010. He is a member of the 2014 IEEE Fellow Committee. Prof. Benediktsson is a co-founder of the biomedical start up company Oxymap (www.oxymap.com). He is a Fellow of the IEEE and a Fellow of SPIE. He received the Stevan J. Kristof Award from Purdue University in 1991 as outstanding graduate student in remote sensing. In 1997, Dr. Benediktsson was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award, in 2000, he was granted the IEEE Third Millennium Medal, in 2004, he was a co-recipient of the University of Iceland's Technology Innovation Award, in 2006 he received the yearly research award from the Engineering Research Institute of the University of Iceland, and in 2007, he received the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society. He was co-recipient of the 2012 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Paper Award and in 2013 he was co-recipient of the IEEE GRSS Highest Impact Paper Award. In 2013 he received the IEEE/VFI Electrical Engineer of the Year Award. He is a member of the Association of Chartered Engineers in Iceland (VFI), Societas Scinetiarum Islandica and Tau Beta Pi.

Lorenzo Bruzzone (S'95-M'98-SM'03-F'10) received the Laurea (M.S.) degree in electronic engineering (summa cum laude) and the Ph.D. degree in telecommunications from the University of Genoa, Italy, in 1993 and 1998, respectively. He is currently a Full Professor of telecommunications at the University of Trento, Italy, where he teaches remote sensing, radar, pattern recognition, and electrical communications. Dr. Bruzzone is the founder and the director of the Remote Sensing Laboratory in the Department of Information Engineering and Computer Science, University of Trento. His current research interests are in the areas of remote sensing, radar and SAR, signal processing, and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects. Among the others, he is the Principal Investigator of the *Radar for icy Moon exploration (RIME)* instrument in the framework of the *JUpiter ICy moons Explorer (JUICE)* mission of the European Space Agency. He is the author (or coauthor) of 150 scientific publications in referred international journals (101 in IEEE journals), more than 215 papers in conference proceedings, and 16 book chapters. He is editor/co-editor of 11 books/conference proceedings and 1 scientific book. His papers are highly cited, as proven form the total number of citations (more than 9850) and the value of the h-index (51) (source: Google Scholar). He was invited as keynote speaker in 24 international conferences and workshops. Since 2009 he is a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society. Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (Seattle, July 1998). Since that time he was recipient of many international and national honors and awards. Dr. Bruzzone was a Guest Co-Editor of different Special Issues of international journals. He is the co-founder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (MultiTemp) series and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003 he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. Since 2013 he has been the founder Editor-in-Chief of the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE. Currently he is an Associate Editor for the IEEE TRANSACTION ON GEOSCIENCE AND REMOTE SENSING and the CANADIAN JOURNAL OF REMOTE SENSING. Since 2012 he has been appointed Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society.