Copyright © 2010 IEEE

Reprinted from

Ferdinando Giacco, Christian Thiel, Luca Pugliese, Silvia Scarpetta, and Maria Marinaro. Uncertainty Analysis for the Classification of Multi-Spectral Satellite Images Using SVMs and SOMs. *IEEE Transactions on Geoscience and Remote Sensing*, volume 48, issue 10, pages 3769-3779, 2010.

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of Universität Ulm's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

Uncertainty Analysis for the Classification of Multispectral Satellite Images Using SVMs and SOMs

Ferdinando Giacco, Christian Thiel, Luca Pugliese, Silvia Scarpetta, and Maria Marinaro

Abstract—Classification of multispectral remotely sensed data with textural features is investigated with a special focus on uncertainty analysis in the produced land-cover maps. Much effort has already been directed into the research of satisfactory accuracyassessment techniques in image classification, but a common approach is not yet universally adopted. We look at the relationship between hard accuracy and the uncertainty on the produced answers, introducing two measures based on maximum probability and α quadratic entropy. Their impact differs depending on the type of classifier. In this paper, we deal with two different classification strategies, based on support vector machines (SVMs) and Kohonen's self-organizing maps (SOMs), both suitably modified to give soft answers. Once the multiclass probability answer vector is available for each pixel in the image, we studied the behavior of the overall classification accuracy as a function of the uncertainty associated with each vector, given a hard-labeled test set. The experimental results show that the SVM with one-versus-one architecture and linear kernel clearly outperforms the other supervised approaches in terms of overall accuracy. On the other hand, our analysis reveals that the proposed SOM-based classifier, despite its unsupervised learning procedure, is able to provide soft answers which are the best candidates for a fusion with supervised results.

Index Terms—Land-cover maps, remotely sensed images, self-organizing maps (SOMs), soft classification, support vector machines (SVMs), uncertainty.

I. INTRODUCTION

THE dimensionality, the amount, and the heterogeneity of remotely sensed data available today require advanced and innovative techniques to extract information and thematic maps useful for environmental monitoring. In the last years, new methods based on optimization and neural network algorithms have been proposed [1], and among them, support vector machines (SVMs) and self-organizing maps (SOMs) are

Manuscript received August 6, 2009; revised December 28, 2009 and February 3, 2010. Date of publication May 27, 2010; date of current version September 24, 2010.

- F. Giacco and S. Scarpetta are with the Department of Physics, University of Salerno, 84084 Salerno, Italy and also with National Institute for Nuclear Physics (INFN), Gruppo Collegato di Salerno, Sezione di Napoli, Italy (e-mail: giacco@sa.infn.it).
- C. Thiel was with the Institute of Neural Information Processing, University of Ulm, 89069 Ulm, Germany. He is now with Crealytics GmbH, 94032 Passau, Germany.
- L. Pugliese is with the International Institute for Advanced Scientific Studies "Eduardo R. Caianiello," 84019 Vietri sul Mare, Italy.
- M. Marinaro, deceased, was with the Department of Physics, University of Salerno, 84084 Salerno, Italy, and also with the International Institute for Advanced Scientific Studies "Eduardo R. Caianiello," 84019 Vietri sul Mare, Italy.
- Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2010.2047863

very promising [2]-[4]. They are particularly useful for highdimensional and multisource data analysis, which is generally difficult to accomplish with classical statistical methods. The freedom from assumptions about the form and distribution of input data makes neural networks a suitable tool to treat both spectral and spatial (texture and context) features, which is a key point in classification of medium-high-resolution images. Spatial information was found to considerably improve the classification ability in many problems, when the spatial scale of the texture is properly chosen [5]. This study focuses on multiclass land-cover classification of multispectral and multisource remotely sensed images with different spatial resolutions: A high-resolution image registered by the IKONOS sensor (4 m/pixel) and a low-medium-resolution image registered by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor (15 m/pixel) have been used. To exploit the high-resolution images, we extracted some textural features using the gray-level co-occurrence matrix (GLCM) [6], [7] and merged them with spectral information taken from the medium-resolution images. We compared the aforementioned classification methods, trained on hard-labeled data, and studied their performances in terms of overall accuracy on hardlabeled data. However, in many situations, and particularly in image classification, classes are often overlapping, mixed, or fuzzy and much uncertainty is associated with the meaning and interpretation of the final land-cover map. Therefore, the wish is to have a soft classifier that is able to provide, for a given pattern vector, an estimation of the membership degrees to the different investigated classes. Therefore, we set the SVMs and SOMs to give soft answers.

Early works using SVMs showed encouraging results [3], [8], and comparative studies stated that classification by an SVM can be more accurate than popular contemporary techniques such as multilayer neural networks or decision trees, as well as conventional probabilistic classifiers such as the maximum likelihood classifier [4], [9], [10]. Although these experiments point out the generalization capability of SVMs, until now, they are mainly used as hard classifiers and their performances were not satisfactorily investigated in terms of uncertainty analysis. With the SVM being a binary classifier, we considered its extension to multiclass architectures following [11] and [12], where the authors deal with producing soft answers. In this paper, two multiclass SVM architectures are considered: the one-versus-one (1vs1), where l(l-1)/2 binary classifiers are combined (one for each pair of classes, where l is the number of classes), and One versus All (1vsALL), where lbinary classifiers are applied on each class versus the others. In particular, we focus on the 1vs1 case where no ready procedure exists, using the technique of pairwise coupling, based on the statistical Bradley–Terry model [13].

SOMs also proved to be very useful in many remote sensing applications [2], [14]-[16]. The basic idea in using an unsupervised strategy, such as the SOM, to achieve classification tasks is to reduce the underlying dependence on the given input and output provided by the labeled training data set. This is a fundamental issue in remote sensing applications, where the failure to exhaustively define classes can result in substantial errors which may also pass undetected in the assessment of classification accuracy [17], [18]. After the unsupervised SOM training, a further step is necessary in which the output SOM clustering is allowed to learn class labels. The preformed clusters in the SOM may aid in accurate and detailed classification, by helping to prevent the learning of inconsistent class labels. Many experiments have been conducted on the application of the SOM for hard classification [1], [2], [19]–[21]. We propose two strategies to use SOM as a soft classifier, taking into account the weighted distance information between pixels and prototypes to get soft outputs. This method gives promising results and reduces, at the minimum, the user's (supervised) influence to extract class information from the unsupervised map. The comparison among the experimental results is then carried out by introducing uncertainty measures on the soft answers produced by SVMs and SOMs. Interestingly, in this scenario, we found that the output of a single classifier can be exploited at its best if one looks at the distribution of the uncertainties. Moreover, the uncertainty analysis allows us to draw up a fusion scheme for the answers obtained from the different classifiers, whose result will exhibit a further improvement in the overall accuracy computed on the hard-labeled test set.

The remainder of this paper is organized as follows. Section II describes multispectral data available and the techniques to extract spatial information from spectral data. Section III presents the SVMs in a multiclass setting. Section IV describes the SOM-based classification strategy with special attention to the fuzzification method used to obtain soft-output answers. The uncertainty analysis is discussed in Section V, where the differences between the proposed classifiers are analyzed in detail through uncertainty measures. Then, in order to fully exploit the experimental results, a simple fusion scheme is described in Section VI, while Section VII gives a brief summary along with some conclusions.

II. DATA SETS AND SPATIAL-FEATURE EXTRACTION

Our study focuses on multispectral ASTER images of two different regions in the province of Salerno (southern Italy). The spectral data of both regions are described in the following two sections (Sections II-A and B), while only for the first case study, we used an IKONOS image in order to add spatial features to the spectral ones (Section II-C).

A. Data Set I Description

The first area of interest is a coastal plain in the southern part of the province of Salerno. Land use is primarily agricultural, but during the last 60 years, an urbanization phenomenon occurred, giving rise to a very indented and complex landscape.

Consequently, the principal types of land covers are agricultural fields (both fallow fields and crop-covered ones), rural fabrics (greenhouses), sea water, a coniferous wood strip along the coastline, and small urban areas made up of discontinuous fabric mixed with vegetation. Two types of multispectral satellite data [21] have been considered in this paper.

- Images captured by the ASTER on the National Aeronautics and Space Administration's Terra satellite [22]. The
 ASTER sensor collects data in 14 bands, going from visible to thermal regions of the electromagnetic spectrum.
- 2) Data recorded by IKONOS 2, a commercial Earthobservation satellite, which offers high-resolution images [23]. The data are collected in four multispectral bands (from visible to near-infrared regions), plus a panchromatic band, namely, a black-and-white image sensitive to all visible radiations.

In this paper, we used the first nine ASTER bands (from visible to short infrared wavelengths) as spectral information and three IKONOS bands to extract spatial features. A first preprocessing stage on both ASTER and IKONOS images has been performed "in-house." In particular, the ASTER data are a Level A product [24], providing band-to-band registration (within and between the sensors) and projection to a common map system, UTM/WGS84. The IKONOS data are a Geo product, which is geometrically corrected (Earth curvature and Earth rotation effect) and also projected to the map system UTM/WGS84 [25]. The choice of the in-house preprocessing levels has been made considering the favorable orography structure of the area of interest and the scope of our investigation.

A second preprocessing stage has been performed with the geographical-information-system software IDRISI Andes, produced by Clark Labs of Clark University. Through the software IDRISI, we first verified the coregistration and georeferencing of both ASTER and IKONOS images. Then, we used the module EXPAND in order to work with image pixels of the same size. The EXPAND module, indeed, simply resizes a pixel, without changing its original value. Namely, six ASTER bands (from four to nine) were resized from 30 to 15 m/pixel, and the IKONOS bands were resized from 4 to 1 m/pixel. The expansion of the IKONOS bands, as described in the next paragraph, will allow us to associate spatial information computed on a window of 15×15 pixels (at 1 m of resolution) directly to one ASTER pixel (at 15 m of resolution).

A further step toward the experimental setup has been to provide a labeled data set. Out of the 236 985 image pixels, expert photointerpreters labeled two spatially separated sets of pixels with their correct land-cover class. Due to their a priori information on the territorial variability, they focused on seven land-cover classes: vegetated land (class 1), built-up area (2), pine wood (3), urban green (4), greenhouses (5), not-vegetated land (6), and water (7). We asked the photointerpreter to find pixels mainly occupied by only one class in order to have a hard-labeled data set. Once a land cover was identified inside the image, a certain number of pixels were labeled with the corresponding class. Then, for each class, the selected pixels were divided into a training and a test set according to the general rule that neighboring pixels should be included in the same labeled subset. The numerical composition of the labeled set is shown in Table I, while Fig. 1 shows a so-called ASTER false-color image where the natural red-green-blue colors are

TABLE I DATA SET I: NUMBER OF LABELED SAMPLES. THE INVESTIGATED LAND COVER CLASSES ARE AS FOLLOWS: (1) VEGETATED LAND, (2) BUILT-UP AREA, (3) PINE WOOD, (4) URBAN GREEN, (5) GREENHOUSE,

(6) NOT-VEGETATED LAND, AND (7) WATER

Class number	1	2	3	4	5	6	7	Sum
Labelled set	249	188	226	251	273	233	238	1657
Training set	145	90	162	159	166	144	163	1029
Test set	104	98	64	92	106	89	75	628



Fig. 1. ASTER false-color composition of the first area of interest (Data Set I). The colored points in the photograph indicate the pixels selected for the training set. The legend is as follows: (Brown) "Vegetated land" (class 1), (black) "builtup area" (class 2), (orange) "pine wood" (class 3), (red) "urban green" (class 4), (blue) "greenhouses" (class 5), (green) "not-vegetated land" (class 6), and (yellow) "water" (class 7).

represented by bands 2 (wavelength of 0.63–0.69 μ m), 1 (wavelength of 0.52–0.60 μ m), and 1 (wavelength 0.52–0.60 μ m), respectively. Fig. 1 also shows differently colored points indicating the training-set pixels.

B. Data Set II Description

The second data set is an ASTER acquisition of the internal (far-from-the-sea) region in the neighborhood of the small city of Battipaglia (Province of Salerno). The landscape complexity is almost the same as that of the previous image. We focused on the detection of four land-cover classes: vegetated land (class 1), not-vegetated land (class 2), greenhouses (class 3), and built-up area (class 4). In this case, the classification task is facilitated, on the one hand, because the number of classes is reduced, but is complicated, on the other hand, because we are not using any kind of spatial features. The image is made up of 150 801 pixels, while the labeled data set (provided with the same techniques previously described) is reported in Table II. Dealing with ASTER data, the whole preprocessing stage is unchanged, i.e., each image pixel is associated with a 9-D vector representing an area on the ground of 15 m².

C. Spatial Preprocessing

Textural features extracted from IKONOS images were introduced. This was done in order to add intrapixel spatial information to the ASTER spectral data of Data Set I. The

TABLE II DATA SET II: NUMBER OF LABELED SAMPLES. THE INVESTIGATED LAND-COVER CLASSES ARE AS FOLLOWS: (1) VEGETATED LAND. (2) NOT-VEGETATED LAND, (3) GREENHOUSE, AND (4) BUILT-UP AREA

Class number	1	2	3	4	Sum
Labelled set	245	309	312	191	1057
Training set	98	139	158	92	487
Test set	147	170	154	99	570

textural features were computed on two different IKONOS data sets: the panchromatic band and the band ratio between nearinfrared and red (4-m/pixel resolution resized to 1 m/pixel), which, in remote sensing literature, is considered as a way to emphasize vegetation [1]. The spatial features were obtained from the well-known GLCM, which is widely used in landcover mapping [5], [26], [27]. A moving window of 15×15 IKONOS pixels (1-m/pixel resolution) is used in the computation of the GLCM, since a window of such dimensions covers the same spatial area as one ASTER pixel. In the computation of the GLCM, data are typically scaled to some fairly modest range of integers (for example, 0-7 in this work, such that the GLCM is a 8×8 matrix). After the GLCM is generated for each direction (horizontal, vertical, left diagonal, and right diagonal), the statistical measures are extracted, and then, the four directions are averaged to remove directional effects; this last choice is due to the absence of preferred directions in the geometry of the investigated land-cover classes. Among the several statistical measures which can be extracted from the GLCM to describe specific textural characteristics of the image [28], we chose the following two: the correlation function computed on the IKONOS panchromatic band and the energy function computed on the IKONOS band ratio. The explicit form of the aforementioned functions is

$$COR = \sum_{i}^{N_G} \sum_{j}^{N_G} \frac{Cov(i, j)}{Stdev(i)Stdev(j)}$$

$$ENE = \sum_{i}^{N_G} \sum_{j}^{N_G} |p(i, j)|^2$$
(2)

$$ENE = \sum_{i}^{N_G} \sum_{j}^{N_G} |p(i,j)|^2$$
 (2)

where i and j are the row and column indexes, N_G is the total number of gray levels, and p(i, j) is the element of the normalized GLCM, while Cov and Stdev are the covariance and the standard deviation, respectively. Our previous analysis showed that these particular choices of statistical measures provide the best classification performances on this data set. Summing up, our data vectors are made up of 11 components, the first nine standing for the spectral information (taken from ASTER bands) and the last two representing textural measures extracted from IKONOS images. The scheme of the preprocessing stage for Data Set I is also shown in Fig. 2.

III. MULTICLASS SVMS WITH SOFT ANSWERS

SVMs were originally developed for the discrimination of two-class problems [29]. They have recently become a popular method in pattern classification for their ability to cope with small training sets and high-dimensional data [30], [33]. A fundamental advantage of the SVM approach is that it facilitates the separation of the class samples, by mapping the input space into a higher dimensional dot-product Hilbert space,

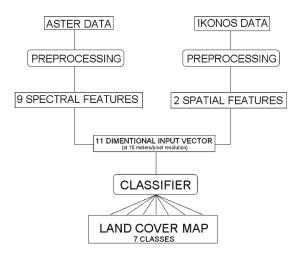


Fig. 2. Scheme of the preprocessing stage used in the classification of Data Set I. The two sets of images, ASTER and IKONOS, are separately preprocessed and coregistered in a common reference system. Then, two textural measures are extracted from the GLCM computed on two different IKONOS bands in order to add spatial features to the nine ASTER spectral bands.

using so-called Mercer kernels. The beauty here is that the transformation does not need to be calculated explicitly and expensively, as distances in the Hilbert space can be computed solely using the kernel.

Regarding the possibilities to extend the originally binary SVMs to multiclass settings, there has been quite some research recently [31], and architectures like 1vs1 and 1vsALL are widely used nowadays (see [32]–[34] for comparisons). The 1vs1 and 1vsALL SVM architectures are widely used in remote sensing applications [35]–[37]. Here, we pay special attention to the possibility of extending them to multiclass soft answers [38]. In a recent work [12], the authors dealt with the issue of accepting, and more importantly producing, soft labels in multiclass SVMs. In the following, we will briefly present the solutions we decided to explore in the current application and elaborate a bit on the 1vs1 case where no ready procedure existed.

In the current case, we have samples from L=7 different classes for Data Set I and L=4 for Data Set II. The 1vsALL approach builds L different SVMs, each of which is able to separate one specific class from all the others. Presented with a new sample x, each SVM $_i$ will answer with the distance $d_i(x)$, $i=1,\ldots,L$, that this sample has to the separating hyperplane. To transform these distances to soft-output answers o_i , we used a sigmoid function (proposed by Platt [38])

$$o_i(d_i(x)) = 1/(1 + \exp(-A_i d_i(x) + B_i)), \qquad i = 1, \dots, L.$$
(3)

The parameters A_i , $B_i \in \mathbf{R}$ are estimated for each SVM_i to minimize the mean square error on the training data between the original label and the sigmoid output, using a batch gradient descent technique (see also [12]). Given that the convergence of this method is problem dependent, we refer to the work of Lin *et al.* [39] for details about the implementation of more robust algorithms.

The solution is not so straightforward in the 1vs1 approach. Here, an SVM $_i$ is built for every pair of classifiers (resulting in L(L-1)/2 machines). To get the desired L-dimensional soft output, the technique employed in most cases today, for exam-

ple, in [40], is as follows: Using an indicator function, transform each of the answers d_i into a vote for one of the two classes distinguished by the current machine i. Then, sum those votes per class and normalize the resulting soft label. This method does not have a bad performance, but has some limitations as it does not deliberately take into account the distance information provided by the values d_i . To heal this issue, one can proceed similarly to the 1vsALL case and use a sigmoid, with the same parameters for all pairs, to transfer the distances to soft answers, which can then be summed up and normalized. However, even then, the class-pair information provided by the SVM_i is not used. In literature, the problem of multiclass probability estimates starting from class-pair information is addressed by different techniques. In our work, we used the method of Hastie and Tibshirani [13], based on the statistical Bradley-Terry model. It uses initial estimations for the pairwise probabilities and, in an iterative process, produces soft labels that take into account the coupled distance information. An introduction to the existing techniques and many interesting theoretical conclusions on pairwise coupling can be found in [41].

Our experimental results were evaluated, for different SVM architectures, in terms of the overall accuracy computed on the test set. In the first case study (Data Set I), the best 1vs1 performance is 95.4%, obtained with a linear kernel, while the maximum for the 1vsALL architecture is 91.8%, with a radial basis function (RBF) kernel. More generally, polynomial and RBF kernels reached at least the same accuracy level as that of the linear one, but are more expensive to calculate. The optimal value for the SVM slack parameter C was chosen out of the range of $[10^{-4}, \ldots, 10^4]$. For Data Set II, the best result is 90.7% for the 1vs1 SVM and 91.2% for the 1vsALL, both with a linear kernel. It must be stressed that the performance of SVMs is highly dependent on selecting the appropriate kernel and its parameters for each data set.

IV. SOM-BASED CLASSIFICATION STRATEGY

As is well known, the SOM is an unsupervised algorithm which achieves two goals: 1) a clustering of the input data into nodes and 2) a local spatial ordering of the map in the sense that the prototypes are ordered on the lattice such that similar inputs belong to topographically close nodes. Such an ordering of the data makes the SOM a powerful clustering algorithm and facilitates the understanding of data structures. Since SOMs have properties of both vector quantization and vector projection, they have been used both for unsupervised applications and as classifiers. When used as a classifier, a two-step strategy is needed, which makes use of unlabeled data to train the SOM and uses a limited number of hard-labeled data (training set) to associate the nodes with class memberships.

The quality of labeled training data can be a source of problems in remote sensing applications, because of the presence of mixed pixels, correlations among training patterns taken from the same area, and so on. Moreover, an exhaustive definition of the classes cannot be easily faced in a supervised net, which is built using only hard labels.

On this basis, the SOM algorithm has some advantages over supervised strategies because the results are created based on the clustering structure, which depends on the entire unlabeled data set used to train the map. Although many experiments have been conducted on the application of the SOM for hard classification [19]–[21], few works have addressed the issue of SOM for soft classification [42].

Usually, a majority vote technique is used to associate a label with a SOM node: Each labeled pixel associated with a node provides a vote for one of the classes, and the class having the majority of the votes is associated with that node. Moreover, each vote is weighted by a frequency factor $1/N_l$, where N_l is the total number of samples of class l in the training set. In spite of its not bad performance, this method does not take into account the distance information between the labeled pixels and the prototype, providing the same label to all the pixels associated with one node. Soft answers are not available in this way. Li and Eastmann [42] get through this problem by associating each SOM output node with a soft answer according to the number of training pixels belonging to the node. Namely, a node k is associated with a soft output, such that its membership to class l is given by

$$P_k^l = \frac{\sum_{j \in S(l,k)} w_j^l}{\sum_{l=1}^L \sum_{j \in S(l,k)} w_j^l}$$
(4)

where S(l,k) is the set of indexes of pixels with label l in node k and

$$w_i^l = 1/N_l \tag{5}$$

is the inverse of the total number of pixels of class l in the training set. This provides a measure which is a sort of posterior probability of the node to belong to a given class l. The soft classification of the pixels is then achieved by assigning the membership values of the node to all pixels which fall into it. This method, hereafter called Fuzzy SOM M1, still has some limitations because it gives the same soft answer to all the pixels associated with one node and it does not take into account the distance information between the labeled pixels and the prototype of each node.

Hence, we propose a different technique, Fuzzy SOM M2, which considers the distance between the pixels associated with a node and all the labeled pixels associated to that node. In this way, we get, for each pixel, a soft label which depends on the position of the pixel inside the node it belongs to. The strategy is as follows: The soft label to be associated with a test pixel \mathbf{y} , which falls into node k, is a weighted mean of the labels of all labeled pixels associated with that node k, where the weights depends on the distances, i.e.,

$$P_k^l(\mathbf{y}) = \frac{\sum_{j \in S(l,k)} w_j^l(\mathbf{y})}{\sum_{l=1}^L \sum_{j \in S(l,k)} w_j^l(\mathbf{y})}.$$
 (6)

S(l,k) again is the set of pixel indexes with label l in node k, and the weights $w_i^l(\mathbf{y})$, now depending on the sample y, are

$$w_j^l(\mathbf{y}) = \frac{1}{N_l} \exp\left(-\frac{\|\mathbf{x}_j^l - \mathbf{y}\|^2}{2\sigma_k}\right), \quad j \in S(l, k)$$
 (7)

where \mathbf{x}_j^l is the jth training pixel of class l and the spread σ^k was set for each node as the mean-square displacement between the prototype k and labeled pixels associated with that

prototype. The difference with respect to (4) (see [42]) is that, here, the weights take into account not only the total number of pixels of class l in the training set but also the distances between the pixel y and the labeled pixels of the training set associated with the node into which the test pixel falls. This method, based on weighted voting instead of simple voting, is, in some sense, analogous to the strategy we used in the multiclass soft SVMs in the previous section, i.e., it is a weighted average of answers, where the weights depend on some distance information. As usual, when the SOM is used as a classifier, it may happen that some nodes are unlabeled, in the sense that there are no training pixels falling into them. In our procedure, if the test pixel falls in a node without training pixels, then it will be considered unclassified. The presence of unclassified nodes means that the L classes are not exhaustive and that there is information on the ground that is not contained in the labeled set.

The two SOM-based classification approaches previously described were first used for a hard classification task. The resulting overall accuracy on Data Set I, computed on the hard-labeled test set of Table I, is 93.3% for the first fuzzification method (Fuzzy SOM M1) and 92.7% for the second one (Fuzzy SOM M2). On the second data set, both methods provided an accuracy of 89.2%. A better understanding of the differences between these two methods will be made possible in the next section, where uncertainty distribution analysis will show that the soft answers provided by the weighted method have to be preferred.

V. UNCERTAINTY ANALYSIS

Understanding and recognizing the uncertainty in image classification and the desire to fully exploit the information content of the produced land-cover maps were the driving forces in the development of soft classification of remotely sensed data [18], [43]. In this paper, we compare the proposed classification methods, described in Sections II and III, by looking at their soft answers on the test set, as well as on the remaining (not labeled) data set.

The pixel's multi-answer output $o \in [0, 1]^L$ (for L classes), yielded by classification, reflects the differences in uncertainty in the resulting classification. It may be considered indicative of dubious classifications, of mixed pixels, of heterogeneous classes, or of fuzzy boundaries between classes [44].

Various measures of uncertainty exist in the literature. In this paper, we consider two of them: The first is based on the maximum probability appearing in the output probability vector associated with each image pixel. This value is an expression of the strength of the class assignment and of possible confusion with other classes.

The second approach is the α quadratic entropy, which was first used in theoretical physics by Fermi and widely applied in risk evaluation of the nearest neighbor classification rule [45]. This measure is based on the concept of the multiplicative class introduced by Pal and Bezdek [46] and is explicitly given by

$$H_{\alpha}(o) = \frac{1}{L2^{-2\alpha}} \sum_{d=1}^{L} o_d^{\alpha} (1 - o_d)^{\alpha}$$
 (8)

where o is a vector representing the soft answers associated with a given pixel, L is the number of the investigated classes, and α

is an exponent which determines the behavior of the uncertainty measure. Indeed, if α is close to zero, the measure is not very sensitive to small changes in the components o_d , while for α close to one, the uncertainty is higher for o_d close to 0.5.

In order to analyze how the uncertainty is distributed among the land-cover classes under consideration for the proposed classifiers, we considered the following strategy: We get, from the soft outputs, an estimation of the measure of uncertainty for a given pixel, and we reject the "hard" allocation when this measure is larger than a given threshold. Hence, in a rejection setting, a classifier is allowed to reject a test sample presented to him, i.e., to refuse to take a decision about the class of the sample. An answer which the classifier says it is not sure about, according to one of the aforementioned uncertainty measures, is rejected. In classical machine learning settings, this is done to increase the classification accuracy on the not-rejected samples, but here, this procedure is used to compare the performances of the different soft classifiers.

Formally, we analyzed the uncertainty of the classifiers on a test set T consisting of M testing samples z in the feature space \mathbb{R}^N (N=11 for Data Set I, and N=9 for Data Set II), with associated labels l that detail to which of the L classes the sample belongs

$$T = \{ (z_{\mu}, l_{\mu}) \mid \mu = 1, \dots, M, \qquad l_{\mu} \in \{1, \dots, L\} \}.$$
 (9)

The answer of the classifier is a soft answer $o \in [0,1]^L$, $\sum_{d=1}^L o_d = 1$, that reflects to what degree the classifier thinks the sample belongs to each class.

The rejection method we decided to use is rather basic

reject if
$$(F(o) > \text{threshold})$$
 (10)

where F is the uncertainty measure for the testing sample o. The F measure is then given by

$$F_{\mathrm{MP}}(o) = 1 - \max_{d} o_d$$
 for the Maximum Probability (11)

or

$$F_{\alpha O}(o) = H_{\alpha}(o)$$
 for the α Quadratic Entropy (12)

where $H_{\alpha}(o)$ is given by (8). The performance on the not-rejected part of the test set, i.e., the number of not-rejected test pixels correctly classified over the total number of not-rejected pixels, is studied as a function of a decreasing uncertainty threshold.

A. Data Set I

In Fig. 3, the overall accuracies for several classifiers are reported as a function of the rejection rate, with rejection based on the maximum probability measure $F_{\rm MP}$. As the rejection rate is not a direct parameter of the algorithm, we simply decreased the rejection threshold from 1 to 0 in steps of 0.01 and noted the respective rates. The figure shows that the 1vs1 Linear SVM steadily provides the highest accuracy on the test set. The 1vsALL SVM approach, even if it starts from lower initial accuracy, has a similar increasing behavior. Concerning the SOM performances, although their initial accuracies (at zero rejection rate) are undoubtedly good, the increasing rejection

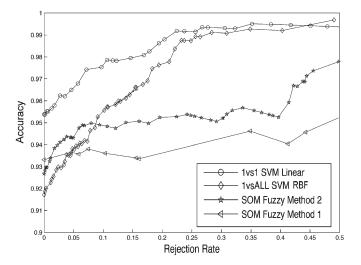


Fig. 3. Data Set I: Overall accuracy, computed on the not-rejected pixels of the test set, as a function of rejection rate when F_{MP} is used. The curves show the results obtained for two SVM- and SOM-based strategies described in the text.

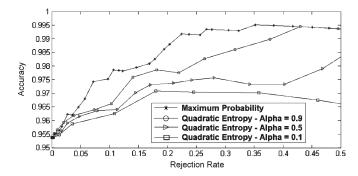


Fig. 4. Data Set I: The 1vs1 Linear SVM overall accuracy as a function of the rejection rate. The curves show the fraction of correctly classified test set pixels, with the rejection based on different uncertainty measures: maximum probability and α quadratic entropy.

does not produce large improvements in the accuracy on the test set. Moreover, it is straightforward to notice that the weighted criterion (Fuzzy SOM M2) to assign the soft labels in the SOM-based strategy has to be preferred with respect to the one which does not take into account the distances (Fuzzy SOM M1). The same analysis has been performed using the second uncertainty measure $F_{\alpha Q}$, where the rejection is based on the α quadratic entropy. The curves obtained show a similar behavior as the ones based on the maximum probability. The only classifier presenting a major difference is the 1vs1 Linear SVM, whose accuracy as a function of the rejection rate is shown in Fig. 4, where different choices for the parameter α are proposed. Quite differently, Fig. 5 shows that the Fuzzy SOM M2 answers, as well as the SVM 1vsALL ones (not reported in this paper), are not sensitive to the choice of the uncertainty measure.

Another investigation is undertaken to evaluate the uncertainty in the produced land-cover maps. To this end, Fig. 6 shows the fraction of retained pixels (class by class) of the whole image with decreasing rejection threshold $F_{\rm MP}$ for three different classifiers: 1vs1 Linear SVM and 1vsALL RBF SVM-and SOM-based classifiers. These curves give an indication of how the uncertainty is distributed among the investigated classes for the two best supervised classifiers and the best Fuzzy-SOM strategy, on the basis of the analysis of Fig. 3.

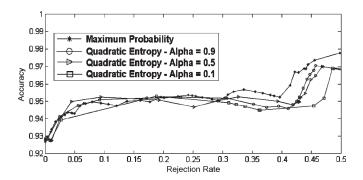


Fig. 5. Data Set I: Comparison among different rejection criteria, for the SOM-based classification experiment (Fuzzy SOM Method 2). In this case, differing from the 1vs1 SVM, the overall accuracy is not influenced by the choice of the uncertainty measure. The accuracies in 1vsALL SVM experiments behave similarly to the SOM ones.

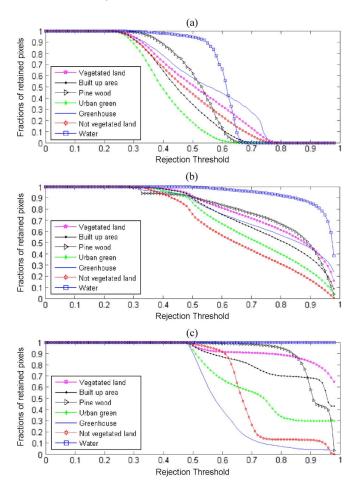


Fig. 6. Data Set I: The curves report the fraction of retained pixels, class by class, as a function of an increasing rejection threshold based on the measure F_{MP} . They refer to the following algorithms: (a) 1vs1 Linear SVM, (b) 1vsALL RBF SVM, and (c) SOM Fuzzy Method 2.

For instance, looking at the curves, we can conclude that the SOM-based classifier assigns a degree of membership of more than 0.99 to the pixels of water, meaning there is practically no uncertainty about this class. From the 1vs1 SVM curves, we can observe that the most easily rejected pixels belong to the class "urban green," and this result fits good to the fact that those pixels are a mix between vegetation and urban structures. On the other hand, the most rejected classes in the

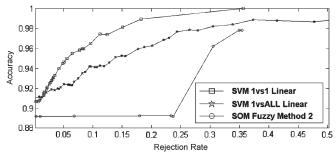


Fig. 7. Data Set II: Overall classification accuracy as a function of rejection rate when $F_{\rm MP}$ is used.

1vsALL SVM and SOM approaches are "not-vegetated land" and "greenhouse," respectively. This result can be attributed, for the most, to difficulties in correctly classifying those two classes. The uncertainty is allocated in a different way for the SOM classification. Fig. 6, indeed, shows that 30% of pixels classified as "urban green" are not rejected even at the maximum threshold, even though it is an informative class difficult to characterize. This is due to some nodes on the SOM output grid which are devoted to mixed pixels and particularly to mixtures of vegetation and urban structures. The SVM rejection curves exhibit a nicely smooth behavior. The 1vs1 SVM answers are much more uncertain (reflected in the early rejection) than the 1vsALL ones, owing to the smoothing effect of the Bradley–Terry coupling procedure that produces the output.

In Section V, we will see how the differences in the proposed approaches can be fully exploited, in order to achieve the best classification accuracy on our test set.

B. Data Set II

The analysis of the second data set gives almost the same results obtained on the first case study. Having reduced the number of input features, the overall accuracies of the classifiers are slightly lower with respect to the previous ones, particularly because of the lack of spatial information. However, the differences in the experimental results between the proposed approaches are essentially the same. In Fig. 7, we show the accuracy of three classifiers as a function of the rejection rate, namely, 1vs1 and 1vsALL SVMs with linear kernel, along with Fuzzy SOM Method 2. Even if the initial accuracy (at zero rejection) of the 1vsALL SVM is higher than those of the other classifiers, the best result at increasing rejection rate is still obtained with the 1vs1 SVM. Fig. 7 refers to a rejection produced by the measure $F_{\rm MP}$, while Fig. 8 shows the impact of the α quadratic measure $F_{\alpha Q}$ on the 1vs1 SVM answers. In this case, as in Data Set I, there is still a dependence on the parameter α even though lower than that in Fig. 4, while the uncertainty measure does not affect at all the performances of the 1vsALL and SOM-based classifiers (such as in Fig. 5). Last, we propose in Fig. 9 the fraction of retained pixels with decreasing rejection threshold based on the F_{MP} measure. The similarity between those curves and the ones in Fig. 6 (Data Set I) is rather plain. Focusing on the curves related to the SOM classification, we found again a different behavior with respect to the SVMs. As Fig. 9 shows, pixels of classes 2 and 4 can be divided in several subgroups, each of which could be associated

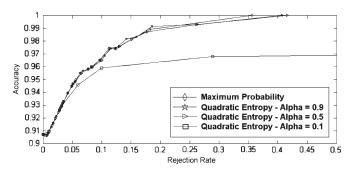


Fig. 8. Data Set II: The 1vs1 Linear SVM overall accuracy as a function of the rejection rate for different uncertainty measures.

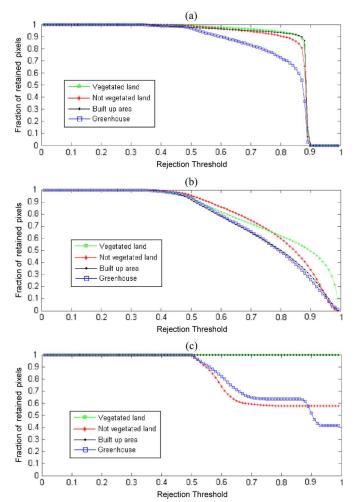


Fig. 9. Data Set II: Fraction of retained pixels, class by class, as a function of an increasing rejection threshold based on the measure $F_{\rm MP}$. The curves refer to classification performed by (a) 1vs1 Linear SVM, (b) 1vsALL Linear SVM, and (c) SOM Fuzzy Method 2.

to a drop in the fraction of retained pixels at different thresholds. This is due, as well as for the classification of Data Set I, to the nature of the SOM output map, where a single class can be associated with several contiguous nodes.

VI. FUSION SCHEME FOR OPTIMAL CLASSIFICATION

In this section, we propose a decision fusion scheme to aggregate the soft outputs of the proposed classifiers, which will emphasize how the uncertainty measure can be useful to

TABLE III

Data Set I: Confusion Matrix for the 1 Versus 1 SVM Classifier With Linear Kernel (Overall Accuracy of 95.4%). Class Names Are as Follows: (1) Vegetated Land, (2) Built-up Area, (3) Pine Wood, (4) Urban Green, (5) Greenhouse, (6) Not-Vegetated Land, and (7) Water

Class	1	2	3	4	5	6	7	Acc. per class
1	104	0	0	0	0	0	0	100%
2	1	80	0	6	8	3	0	81.6 %
3	0	0	64	0	0	0	0	100%
4	0	5	2	85	0	0	0	92.4%
5	0	0	0	0	106	0	0	100%
6	0	0	0	0	4	85	0	95.5%
7	0	0	0	0	0	0	75	100%

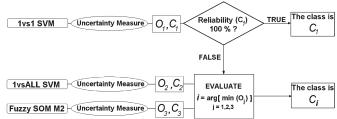


Fig. 10. Fusion scheme used to aggregate the answers of different classifiers. The variables O_i and C_j (j=1,2,3) refer to the uncertainty measure and the class label of a given new sample, respectively. Once checked, the reliability of the best classifier (1vs1 SVM in this case) on class C_1 , the fusion scheme provides an answer which could be C_1 or C_i , where the class C_i is associated with the answer having the minimum uncertainty.

improve the classification results. Given that the best classification performance on Data Set I is achieved by the 1vs1 Linear SVM, we looked at the curves in Fig. 3 and combined the best result with one or both of the other two approaches: the 1vsALL RBF SVM and the best SOM (namely, Fuzzy M2). Moreover, by looking at the confusion matrix of the 1vs1 Linear SVM (see Tables I and III for the class names), we also chose to apply the fusion scheme only to classes 2, 4, and 6, where the performances on the test set are different from 100%.

The rather intuitive combination rule, also shown in Fig. 10, is as follows: If the class assigned by the 1vs1 SVM is a class where no error occurs (reliability, i.e., accuracy per class, is 100%), the final answer is the one provided by the 1vs1 SVM. Otherwise, the sample is also passed to the other classifiers in the architecture, and the class assignment will be fixed by the classifier that provides the least uncertainty in its answer. Considering that the 1vs1 Linear SVM overall accuracy is 95.4%, we report the fusion results obtained with the 1vsALL RBF SVM, Fuzzy SOM M2, and both in Table IV. Evidently, only the fusion of 1vs1 SVM and SOM gives an improvement over the baseline. For comparison, Table V shows the confusion matrix for the aforementioned best combination result, which sports a final accuracy of 96.7%. Another important point, not explicitly reported here, is that if the measure $F_{\alpha Q}$ is used in the fusion scheme, the combination results always have a worse performance than the ones obtained with measure $F_{\rm MP}$.

An interesting result is also found concerning Data Set II. Being a simpler classification task, with only four classes and nine input features, the fusion strategy gives less evident results.

By the way, the validity of the uncertainty analysis is still confirmed in this experiment because we found that the fusion of the 1vs1 SVM with the SOM answers provides a final accuracy, 91.6% (see Tables VI and VII), which is higher than

TABLE IV

DATA SET I: FUSION RESULTS OBTAINED BY COMBINING 1 VERSUS 1 LINEAR SVM, 1 VERSUS ALL RBF SVM, AND FUZZY-SOM CLASSIFIERS

Classifiers	Accuracy
SVM 1vs1 Linear - SVM 1vsALL RBF	95.4%
SVM 1vs1 Linear - Fuzzy SOM M2	96.7%
SVM 1vs1 Linear - SVM 1vsALL RBF - Fuzzy SOM M2	95.4%

TABLE V

DATA SET I: CONFUSION MATRIX CORRESPONDING TO THE BEST FUSION RESULT: 1 VERSUS 1 LINEAR SVM AND FUZZY SOM M2 (OVERALL ACCURACY OF 96.7%). CLASS NAMES ARE GIVEN IN TABLE III

Class	1	2	3	4	5	6	7	Acc. per class
1	104	0	0	0	0	0	0	100%
2	0	82	0	6	5	5	0	83.7 %
3	0	0	64	0	0	0	0	100%
4	0	2	1	89	0	0	0	96.7%
5	0	0	0	0	105	1	0	99.1%
6	0	1	0	0	0	88	0	98.9%
7	0	0	0	0	0	0	75	100%

TABLE VI

Data Set II: Confusion Matrix Corresponding to the 1 Versus 1 Linear SVM Classification. The Overall Accuracy Is 90.7%. Class Names Are as Follows: (1) Vegetated Land, (2) Not-Vegetated Land, (3) Greenhouse, and (4) Built-up Area

Class	1	2	3	4	Acc. per class
1	147	0	0	0	100%
2	0	128	0	42	75.3 %
3	0	0	153	1	99.3%
4	0	10	0	89	89.9%

TABLE VII

Data Set II: Confusion Matrix Corresponding to the Best Fusion Result. The Combination of 1 Versus 1 Linear SVM and Fuzzy SOM M2 Answers Produces a Final Accuracy of 91.6%

Class	1	2	3	4	Acc. per class
1	147	0	0	0	100%
2	0	132	0	38	77.6 %
3	0	0	154	0	100%
4	0	10	0	89	89.9%

the performance of each individual classifier (90.7% for the 1vs1 SVM and 89.2% for the SOM) and slightly higher than one obtained with the 1vsALL classifier (91.2%). On the other hand, the fusion of the 1vsALL soft answers with each of the other classifiers does not generate any improvement. This means that, even though, for this data set, the 1vsALL result is the best classifier from a "hard allocation" point of view, the uncertainty is still best assigned by the 1vs1 SVM. The last outcome is not only evident from the rejection-accuracy plot in Fig. 7 but is also confirmed by the fusion results. By looking at Tables III, V, VI, and VII, we have to remark that, except for class 5 of Data Set I, the fusion scheme always provides a small improvement of the individual class accuracies with respect to the performance of the best single classifier, for both data sets. This improvement, even if small, is not to be taken for granted in a fusion scheme, since the decision rule could also worsen the final accuracy.

VII. CONCLUSION

In this paper, we dealt with the classification of remotely sensed multispectral data in thematic land-cover maps. We im-

plemented several classification strategies based on supervised SVMs, with different architectures, and Kohonen's SOM. We modified the algorithms to obtain soft outputs, and in particular, we focus on the 1vs1 SVM, using the technique of pairwise coupling, based on the Bradley-Terry model, and introduced a new method to get soft outputs from the unsupervised SOM clustering, based on the distances between pixels falling into the same SOM node. In this way, a study of the uncertainty of the hard allocations produced for the land-cover maps is available for all the classifiers considered here. The study of the uncertainty assignment is carried out in a rejection scenario. The results show that the best performance is provided by the 1vs1 SVM with linear kernel. Concerning the choice of the measures used to estimate the uncertainty, we found that, at least for our data set, when applied to the selection of the correct answers, the simple maximum probability yields the best results.

Last, we have shown how the uncertainty measure can be useful in the fusion of several classifiers, showing that, by exploiting the uncertainty information, the decision fusion is able get an improvement over the performance of individual classifiers. The fusion results also show that the SOM algorithm, owing to its unsupervised nature, can be a successful complementary contributor to the quality and accuracy assessment of the final output land-cover map and in the implementation of multiple-classifier systems. Indeed, we found that the best fusion result is not obtained combining the two best classifiers but aggregating the answers of the SOM with one of the supervised SVM.

REFERENCES

- [1] B. Tso and P. M. Mather, Classification Methods for Remotely Sensed Data. New York: Taylor & Francis, 2001.
- [2] T. Villmann, E. Merenyi, and B. Hammer, "Neural maps in remote sensing image analysis," *Neural Netw.*, vol. 16, no. 3/4, pp. 389–403, Apr. 2003.
- [3] G. H. Halldorsson, J. A. Benediktsson, and J. R. Sveinsson, "Support vector machines in multisource classification," in *Proc. Geosci. Remote Sens. Symp.*, 2003, vol. 3, pp. 2054–2056.
- [4] G. M. Foody and A. Mathur, "A relative evaluation of multiclass image classification by support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1335–1343, Jun. 2004.
- [5] M. A. Shaban and O. Dikshit, "Improvement of classification in urban areas by the use of textural features: the case study of Lucknow city, Uttar Pradesh," *Int. J. Remote Sens.*, vol. 22, no. 4, pp. 565–593, Mar. 2001.
- [6] S. E. Franklin and D. R. Peddle, "Classification of SPOT HRV imagery and texture features," *Int. J. Remote Sens.*, vol. 11, no. 3, pp. 551–556, Mar. 1990.
- [7] P. Gong, D. J. Marceau, and P. J. Howarth, "A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data," *Remote Sens. Environ.*, vol. 40, no. 2, pp. 137–151, May 1992.
- [8] J. A. Gualtieri and S. Chettri, "Support vector machines for classification of hyperspectral data," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2000, pp. 813–815.
- [9] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [10] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, no. 4, pp. 725–749, Feb. 2002.
- [11] C. Thiel, F. Giacco, F. Schwenker, and G. Palm, "Comparison of neural classification algorithms applied to land cover mapping," in *Proc. 18th Italian Workshop Neural Netw.*, 2009, vol. 193, pp. 254–263.
- [12] C. Thiel, S. Scherer, and F. Schwenker, "Fuzzy-input fuzzy-output one-against-all support vector machines," in *Proc. 11th Int. Conf. Knowledge-Based Intell. Inf. Eng. Syst.*, 2007, vol. 4694, pp. 156–165.
- [13] T. Hastie and R. Tibshirani, "Classification by pairwise coupling," *Ann. Stat.*, vol. 26, no. 2, pp. 451–471, Apr. 1998.

- [14] W. Wan and D. Fraser, "Multisource data fusion with multiple selforganizing maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1344–1349, May 1999.
- [15] L. Pugliese, S. Scarpetta, A. Esposito, and M. Marinaro, "An application of neural and probabilistic unsupervised methods to environmental factor analysis of multi-spectral images," in *Image Analysis and Processing*. New York: Springer-Verlag, 2005, pp. 1190–1197.
- [16] T. Villmann and E. Merenyi, "Extensions and modifications of the Kohonen-SOM and applications in remote sensing image analysis," in Self-Organizing Neural Networks: Recent Advances and Applications. New York: Springer-Verlag, 2001, pp. 121–144.
- [17] G. M. Foody, "Status of land cover classification accuracy assessment," *Remote Sens. Environ.*, vol. 80, no. 1, pp. 185–201, Apr. 2002.
- [18] G. M. Foody, "Hard and soft classifications by a neural network with a non-exhaustively defined set of classes," *Int. J. Remote Sens.*, vol. 23, no. 18, pp. 3853–3864, Sep. 2002.
- [19] Y. Ito and S. Omatu, "Category classification method using a selforganizing neural network," *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 829– 845, Mar. 1997.
- [20] C. Y. Ji, "Land-use classification of remotely sensed data using Kohonen self-organizing feature map neural network," *Photogramm. Eng. Remote Sens.*, vol. 66, pp. 1451–1460, 2000.
- [21] F. Giacco, L. Pugliese, S. Scarpetta, and M. Marinaro, "Application of neural unsupervised methods to environmental factor analysis of multi-spectral images with spatial information," in *Proc. SPPRA*, 2008, pp. 151–154.
- [22] Y. Yamaguchi, A. B. Kahle, H. Tsu, T. Kawakami, and M. Pniel, "Overview of Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 4, pp. 1062–1071, Jul. 1998.
- [23] G. Dial, H. Bowen, F. Gerlach, J. Grodecki, and R. Oleszczuk, "IKONOS satellites, imagery, and products," *Remote Sens. Environ.*, vol. 88, no. 1/2, pp. 23–36, Nov. 2003.
- [24] H. Fujisada, "ASTER level 1 data processing algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 4, pp. 1101–1112, Jul. 1998.
- [25] E. Baltsavias, M. Pateraki, and L. Zhang, "Radiometric and geometric evaluation of Ikonos GEO images and their use for 3D building modelling," presented at the ISPRS Workshop, Hannover, Germany, 2001.
- [26] R. M. Haralick and L. G. Shapiro, Computer and Robot Vision. Reading, MA: Addison-Wesley, 1992.
- [27] D. J. Marceau, P. J. Howarth, J. M. Dubois, and D. J. Gratton, "Evaluation of the grey-level co-occurrence method for land cover classification using SPOT," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 4, pp. 513–519, Jul. 1990.
- [28] P. M. Mather, Computer Processing of Remotely-Sensed Images. Hoboken, NJ: Wiley, 1999.
- [29] V. N. Vapnik, The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.
- [30] A. R. Webb, *Statistical Pattern Recognition*. Hoboken, NJ: Wiley, 2002.
- [31] C. Angulo, F. J. Ruiz, L. Gonzalez, and J. A. Ortega, "Multi-classification by using Tri-Class SVM," *Neural Process. Lett.*, vol. 23, no. 1, pp. 89– 101, Feb. 2006.
- [32] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.
- [33] B. Schölkopf and A. J. Smola, "Learning with Kernels," in Support Vector Machines, Regularization, Optimization and Beyond. Cambridge, MA: MIT Press, 2002.
- [34] L. Kahsay, F. Schwenker, and G. Palm, Comparison of Multiclass SVM Decomposition Schemes for Visual Object Recognition. New York: Springer-Verlag, 2005, pp. 334–341.
- [35] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Evaluation of kernels for multiclass classification of hyperspectral remote sensing data," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2006, pp. II-813–II-816.
- [36] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.
- [37] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A joint spatial and spectral SVMs classification of panchromatic images," in *Proc. IGARSS*, 2007, pp. 1497–1500.
- [38] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances* in *Large Margin Classifiers*. Cambridge, MA: MIT Press, 1999, pp. 61–74.

- [39] H. T. Lin, C. J. Lin, and R. C. Weng, A note on Platt's probabilistic outputs for support vector machines, Dept. Comput. Sci., Nat. Taiwan Univ., Taipei, Taiwan. [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/papers/plattprob.ps
- [40] A. Gidudu, G. Hulley, and T. Marwala, "Image classification using SVMs: One-against-one vs one-against-all," presented at the 28th Asian Conf. Remote Sens., 2007. [Online]. Available: http://arxiv.org/ ftp/arxiv/papers/0711/0711.2914.pdf. arXiv:0711.2914v1
- [41] T. F. Wu, C. J. Lin, and R. C. Weng, "Probability estimates for multiclass classification by pair wise coupling," *J. Mach. Learn. Res.*, vol. 5, pp. 975–1005, 2004.
- [42] Z. Li and J. R. Eastman, "Commitment and typicality measurements for the self-organizing map," *Int. J. Remote Sens.*, 2010, to be published.
- [43] M. Xu, P. Watanachaturapom, P. K. Varshney, and M. K. Arora, "Decision tree regression for soft classification of remote sensing data," *Remote Sens. Environ.*, vol. 97, no. 3, pp. 322–336, .2005.
- [44] M. F. Goodchild, L. Chih-Chang, and Y. Leung, "Visualizing fuzzy maps," in Visualization in Geographical Information Systems. Hoboken, NJ: Wiley, 1992, pp. 158–167.
- [45] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.
- [46] N. R. Pal and J. C. Bezdek, "Measuring fuzzy uncertainty," *IEEE Trans. Fuzzy Syst.*, vol. 2, no. 2, pp. 107–118, May 1994.



Ferdinando Giacco received the Laurea (M.S.) degree in physics from the University of Naples "Federico II," Napoli, Italy, in 2005 and the Ph.D. degree from the University of Salerno, Salerno, Italy, with a thesis on automatic classification of satellite images.

He is currently holding a postdoctoral position with the University of Salerno. His research interests are in the field of signal processing, pattern recognition, and theoretical physics models for biological applications.



Christian Thiel received the M.Sc. degree (German Diplom) from the Institute of Neural Information Processing, University of Ulm, Ulm, Germany, in 2004, where he worked on several advanced machine-learning-related topics for his dissertation (submitted in December 2009).

He is currently with Crealytics GmbH, Passau, Germany, an early proponent of profit-driven search marketing, where he is in charge of the bid optimization algorithms. His research interests include pattern recognition, particularly multiple-classifier systems,

and working with sparsely labeled data.



Luca Pugliese received the Laurea (M.S.) degree in telecommunication engineering from the University of Naples "Federico II," Napoli, Italy, in 2000 and a Specialization Postdegree Certificate in Applied Informatics at the International Institute for Advanced Scientific Studies (IIAS) "E. R. Caianiello," Vietri sul Mare, Italy, which is a private research and educational institute in 2001.

From 2004 to 2007, he was a Remote Sensing Expert in an interdisciplinary project experience (Innova Project) promoted by Regione Campania

upon EU Funds, concerning satellite remote sensing application to archeological studies. From 2003, he collaborates in educational activities with IIAS. His interests span from geographic information systems to image processing applied to remote sensing data analysis and particularly to aspects regarding the dissemination of related research topics to both private practitioners and public body technicians.



Silvia Scarpetta received the Ph.D. degree in physics from the University of Salerno, Salerno, Italy, with a thesis on online learning in neural networks supervised by Prof. Maria Marinaro, in 1999.

She has been a Permanent Researcher with the Faculty of Engineering, University of Salerno, since 2002. Her scientific interests range from statistical mechanics and neurobiological application of theoretical physics to artificial neural networks and to pattern recognition for environment monitoring.



Maria Marinaro received the Ph.D. degree in physics from the University of Naples "Federico II," Napoli, Italy, in 1959.

She had been a Full Professor in theoretical physics since 1979 with the University of Salerno, Salerno, Italy, where she was the Dean of the Faculty of Science (1976–1978 and 1983–1995). She had been the President of the International Institute for Advanced Scientific Studies "Eduardo R. Caianiello," Vietri sul Mare, Italy, since 1993. Her scientific interests covered many fields of physics,

such as renormalization in quantum field theory, condensed matter, strongly correlated systems, statistical mechanics, cortical neural models, and artificial neural networks and their applications to signal classification and remote sensing images. She passed away in 2009.