Comparison of Computational Intelligence Based Classification Techniques for Remotely Sensed Optical Image Classification

Demetris Stathakis and Athanassios Vasilakos

***Abstract*—Several computational intelligence components, namely neural networks (NNs), fuzzy sets, and genetic algorithms (GAs), have been applied separately or in combination to the process of remotely sensed data classification. By applying computational intelligence, we expect increased accuracy through the use of NNs, optimal NN structure and parameter determination via GAs, and transparency using fuzzy sets is expected. This paper systematically reviews and compares several configurations in the particular context of remote sensing for land cover. In addition, some of the configurations used here, such as NEFCASS and CANFIS, have few previous applications in the field. A comparison of the configurations is achieved by testing the different methods with exactly the same case-study data. A thorough assessment of results is performed by constructing an accuracy matrix for each training and testing data set. The evaluation of different methods is not only based on accuracy but also on compactness, completeness, and consistency. The architecture, produced rule set, and training parameters for the specific classification task are presented. Some comments and directions for future work are given.**

***Index Terms*—Fuzzy neural networks (FNNs), fuzzy sets, genetic algorithms (GAs), neural networks (NNs), remote sensing (RS).**

I. Introduction

NEURAL NETWORKS (NNs) [50], fuzzy-set theory [87], and genetic algorithms (GAs) [38] are the main components of computational intelligence (CI) [32]. Each of these components has been individually applied to the task of satellite image classification. A number of NN applications are available in the remote-sensing (RS) literature. NNs are often reported to yield comparable or superior accuracy levels compared to statistical classifiers. They have also proven to be a suitable paradigm for multisource classification by the incorporation of ancillary information, such as topographic, contextual, etc. [10]. Ongoing research is focused on the determination of the structure of the network as it has always been hard to determine the number of hidden layers as well as the number of nodes per hidden layer [42]. Another area of research is the translation of the trained NN components into meaningful information. In certain applications, such as security and medical applications, black-box NNs are not acceptable [30]. Extracting rules is useful to understand what the NNs have learned and how they predict [31], [80].

Manuscript received September 22, 2005; revised February 13, 2006.

D. Stathakis is with the Joint Research Centre of the European Commission, MARS-FOOD, 21020 (VA), Ispra, Italy (e-mail: demetris.stathakis@jrc.it).

A. Vasilakos is with the Department of Planning and Regional Development, University of Thessaly, Volos 38334, Greece (e-mail: vasilako@ath.forthnet.gr).

Digital Object Identifier 10.1109/TGRS.2006.872903

TABLE I
Kanellopoulos–Wilkinson Rule

	1 hidden layer	2 hidden layers
upper bound	max{M:3M:N, M:3N:N}	M:3M:3N:N
lower bound	max{M:2M:N, M:2N:N}	M:2M:2N:N

M = number of input units
N = number of output units

The rule suggests that optimum topology should be somewhere between the upper and lower bounds.

Many authors characterize an NN as standard when they refer to a feedforward multilayer perceptron (MLP) trained by the backpropagation (BP) algorithm [71], [85]. Given that structure identification is equally important as is parameter identification [39], we extend here the notion of a standard NN to reflect its structure as well. The purpose is that the actual structure of an NN can yield very different results even if the feedforward BP-MLP condition is kept. To standardize the structure, we use the Kanellopoulos–Wilkinson rule [42]. Although this is a heuristic and clearly is not a panacea, it can be used, given the lack of any other concrete rule, to highlight significant deviations from common practice. Letting M be the number of input units and N be the number of output units, the rule is summarized in Table I. To be safe, the first hidden layer can also be set to four times the number of inputs, according to [42].

The fuzzy-set theory has great potential in the context of classifying remotely sensed data, since land use and land cover are definitely one of those concepts that are a matter of degree [23], [83], [89]. The complexity of RS data classification makes fuzzy-set theory a very attractive alternative in capturing the semantics of information as fuzzy if–then rules. The latter are an effective way of presenting summarized information that can be easily perceived by humans. Fuzzy logic (FL) is used as a means of information granulation [88], which is very close to human reasoning for making rational decisions in an environment of imprecision, uncertainty, and partial truth [89]. Fuzzy granulation also opens the way for linguistic processing [14], [15], [67]. There are two main types of fuzzy systems—the Mamdani type and the Takagi–Sugeno (TSK) type [40]. The fuzzy rules in the former type contain linguistic terms as captured by fuzzy sets

$$\text{If } x \text{ is } A \text{ and } y \text{ is } B, \text{ then } z = C \qquad \text{(Mamdani type)}.$$

0196-2892/$20.00 © 2006 IEEE

A, B, and C are fuzzy sets, whereas x and y are the input variables and z is the output variable.

The consequent of the fuzzy rules of the TSK type is a crisp value rather than a linguistic variable, i.e., only the premise is fuzzy. In a first-order TSK-type rule, the consequent is a linear equation of the form.

$$\text{If } x \text{ is } A \text{ and } y \text{ is } B \text{ then } z = px + qy + r \quad \text{(TSK type)}.$$

A and B are fuzzy sets, p, q, and r are constants, whereas x and y are the input variables and z is the output variable.

A GA is a method, inspired by natural evolution, able to perform a randomized global search in a solution space [75]. GAs are characterized by both implicit and explicit parallelism. The latter makes them suitable for parallel computer-architecture implementation much like NNs. The former, implicit parallelism, refers to the unique property of GAs under which it is possible to estimate the average fitness of a much larger number of solutions by implicitly calculating the observed average fitness of solution patterns (schemata) with instances in the actual population of solutions [38], [90, Ch. 4]. It is because of this property that GAs are more likely to converge toward the global solution [22] compared to gradient-descent optimization methods, especially when the solution space to be searched is complex (i.e., not smooth, not unimodal) [90, Ch. 5]. GAs have been applied in RS, mostly as an unsupervised method, because they can automatically determine the number of classes (as opposed to c-means, for example) [3]. They have also been used for parameter estimation in characterizing contextual information [79].

There is a symbiotic combination of CI components for flexible information processing [51]. Recently, this synergy between components has been explored and applied to the classification of remotely sensed data. The synergy is realized in many different architectures [53], [75]. The outline of what we expect in the scope of CI is that accuracy is assured by the use of NNs, the size of the NN is determined by optimization via a GA, and transparency is achieved by the use of fuzzy sets.

The most common synergy, sporadically appearing in the RS literature, is the fuzzy NNs (FNNs) [14]–[16], [66], also known as neuro-fuzzy systems [59] or granular NNs (GNN) [68]. Those are fuzzy systems trained by algorithms originally applied to NNs. The strength of the neuro-fuzzy systems is dual. First, NN training algorithms are quite efficient. Second, fuzzy sets are used to granulate information and fuzzy if–then rules are then derived to convey the semantics of acquired knowledge. The desired outcome of merging NN with FS theory is increased transparency, i.e., the ability to transform the NN skeleton into easy-to-comprehend linguistic rules, rather than increased accuracy [15], [57]. [69]. Logic rules are determined simultaneously with training, eliminating the need for an *a posteriori* search process [29], [31] as in other rule-extraction methods.

One other very interesting synergy is the incorporation of GAs. They are used to automatically determine the optimal NN structure by finding the optimal number of connections (connectivity level) and the number of nodes in the hidden layer [69]. GAs have also been considered as an alternative for training NNs [44]. A training process based on a GA is less likely to be trapped in local maximum and, thus, it is more likely, compared to gradient-descent-based algorithms, to yield a global optimum [72], [73]. It has been proposed that GAs can be used to find a good initial solution and, subsequently, a gradient search optimization, such as BP, can yield improved results [44], [47], [69].

In this paper, we review the different approaches for deploying CI in the classification of remotely sensed imagery to build land-use and land-cover maps. Our objective is to compare and evaluate several different architectures that have been used in this context. A concurrent target is to identify promising aspects of CI methods, as a guideline for future research.

The paper is organized as follows. First, some specific CI architectures are reviewed in Section II. The primary focus is on RS applications. Then, we apply several of these methods to the same data set in order to compare them in Section III. A comparison of experimental results is presented in Section IV and final conclusions in Section V.

In the discussion that follows, we use the abbreviations R, G, B, IR, NIR, and TIR to indicate visible red, visible green, visible blue, infrared, near infrared, and thermal infrared channels, accordingly. All training times given throughout the paper for our own experiments refer to a dual 2.80-GHz processor with 1.5-GB of RAM computer configuration. Input and output vectors used in our experiments are available by e-mail from the authors or, alternatively, by ftp at ftp://agrifish.jrc.it/public/demetris/IEEE/CIinRS/CIinRS.rar

II. Review of CI Architectures in the RS Context

The fuzzy counterpart for most of the standard NN architectures has been proposed. In this section, the applications of some architectures are briefly presented. It has been expressed in [13] that neuro-fuzzy techniques have been mostly applied in pattern recognition and to synthetic data. From reviewing the literature, it seems that, in general terms, this argument still holds today. Remote-sensing applications of CI are not numerous. Since RS for land cover and land use is central to the present paper, we exclude pattern-recognition applications from the discussion as much as possible. It is common in neuro-fuzzy literature to give examples applied to RS (satellite or airborne), real (such as the Iris data set), or synthetic data. Our intention is to include in our research only the first case of data. Nevertheless, the other two data types are very useful as a point of departure for method comparison. Note also that Wilkinson [86] has also excluded simulated satellite data sets as well as results not providing at least one indicator of classification performance with respect to ground reference data.

A. *Fuzzy MLP*

The fuzzy operation on an MLP neuron has been proposed 30 years ago [45]. As with standard MLP, its fuzzy counterpart is also trained with the BP learning algorithm [62], [63]. The assignment of fuzzy output membership values enable the NN to more efficiently classify fuzzy data with overlapping class boundaries [62]. It has been shown that it is possible to generate both positive and negative rules by back tracing along maximal paths using a trained network [52]. Other rule-extraction methods for trained networks could also be used [31]. The fuzzy MLP has been used as an example of the deployment of soft classifiers in pixel spectral unmixing [5]. The architecture used (9:15:4), which is determined by experimental rules [42], is slightly smaller than that of the lower bound derived by the Kanellopoulos–Wilkinson rule. Three fuzzy sets (low, medium, and high) are assigned to each of the three input dimensions. Synthetic data are used to classify four land-cover classes (water, soil, urban, and vegetation).

An effective way of translating any MLP into a rule-based fuzzy system and vice versa is described in [11]. This is achieved by deploying the concept of f-duality and the interactive-OR (I-OR) logical operator. The rules constructed with this less common operator are inevitable slightly less easy to comprehend. This method has been reported to perform poorly in the context of RS data [70].

B. Fuzzy Self-Organizing Map (FSOM)

The FSOM [78], [82] is the fuzzy version of the Kohonen SOM [43]. In [13], an implementation of FSOM receiving two Landsat Thematic Mapper (TM) bands as inputs and classifying into seven land-cover categories is shown. A two-step process is proposed. First, the FSOM is used for unsupervised clustering. Then, a BP-MLP is deployed to label the clusters (supervised step). The number of MLP input nodes is equal to the number of nodes in the SOM output.

In [70], a Gaussian membership function (MF) to fuzzify the traditional SOM is used. A multiband (R, G, IR) aerial image is used to classify into water, wetlands, forest, and built-up. By examining the band histograms, they determined that the optimal number of rules per class is two. Their system significantly outperforms a BP-trained NN and a maximum-likelihood classifier (MaxLike) in terms of overall accuracy. The actual structure of the NN is not given. The derived fuzzy if–then rules are somewhat complex because they are based on the less commonly used I-OR logical operator [11].

The application of an FSOM with a rule-refinement algorithm to two data sets is given in [64]. The first data set is a seven-band Landsat MSS image to be classified into a six-category schema, four of which are soil types, one is vegetation, and one is a mixture of soil and vegetation. The second is a four-band Landsat TM image and the output classes are eight: six land-cover types plus shadow and clouds. In both cases, the classification accuracy achieved is around 85%. The method is proven to yield a relatively compact rule set with less than 30 rules, even when the dimensionality of input is increased from four to seven. The rules contain the "close to" fuzzy set, which is modeled with the Gaussian function. In both cases, the accuracy is given as overall percentages. The number of samples per class is quite uneven. In the second image, two-thirds of the samples belong to a single class, meaning that perhaps an accuracy measure that takes into account the deviation from chance (e.g., k-hat) might yield different results.

C. Fuzzy Adaptive Resonance Theory (ART)

Fuzzy ARTMAP [21] is probably the most used neuro-fuzzy architecture in RS today. Adaptive resonance theory ARTMAP [20] is a supervised method to self-organize arbitrary mapping from input vectors, representing features such as spectral values and terrain variables, to output vectors such as land-cover classes [17], [18].

ARTMAP has been used to classify Landsat TM data into 17 vegetation classes [19]. Apart from the seven spectral bands, the authors have included ancillary information, namely topographic (cosine of the local solar zenith angle, elevation, slope, aspect) and location (Northing and Easting). The method is shown to produce comparable results to an expert system. Interestingly, they find that location alone is not only sufficient to classify the image with comparable accuracy to using only spectral information, but it also results in reducing the rule set by half. In [26], ART-2 (i.e., accepting continuous input) classification of synthetic aperture radar (SAR) data has been performed and some preliminary results compared to statistical and fuzzy c-means methods are reported.

Fuzzy ARTMAP has been found to produce superior results compared to MaxLike in [34], which was tested for global land-cover mapping. Inputs to the system are a year's monthly maximum normalized difference vegetation index (NDVI) values plus geographic location. The outputs are eleven land-cover classes. In addition, Gamba and Dell'Acqua [33] applied fuzzy ARTMAP to classify airborne optical and radar data into three *ad hoc* categories (vegetation, bare soil, and building). They compared the results to an ART configuration. Furthermore, Amici *et al.* [2] deployed fuzzy ART to classify two satellite radar inputs [Earth Resources Satellite (ERS) 1/2 and Radarsat] into four *ad hoc* classes (vegetation, built up, water, and flooded). It is found there that fuzzy ART performs better with the addition of 3×3 contextual information. In another study [48], they used four spectral dimensions [Linear Imaging Self-Scanning II (LISS-II) sensor of the Indian Remote Sensing (IRS)-1B] and 13 output classes, half of them water related. Compared to MaxLike and MLP, fuzzy ARTMAP consistently performs slightly better for the six images used. Fuzzy ARTMAP has also been used to produce a binary mask (coniferous–not coniferous) using three indexes (brightness, greenness, and wetness) as input, derived from spectral Landsat TM [65]. Finally, fuzzy ARTMAP has been used for change detection [1] with respect to 15 desert land-cover classes.

Based on ARTMAP, the fuzzy adaptive learning control network (FALCON) algorithm has also been proposed. Classification is implemented as a three-step process. Fuzzy ARTMAP handles the fuzzy clustering in input and output spaces. GAs are applied to find the fuzzy rules. BP is used to tune the membership functions [22]. Also, the two-stage-hybrid (TSH) method has been introduced. The first stage is an unsupervised learning ART and the second stage is a majority voting mechanism [12]. Baraldi *et al.* [5] used it to estimate the composition of mixed pixels on synthetic data.

Fuzzy ARTMAP is a highly automated process since only three parameters have to be set [21]: the choice parameter ($\alpha > 0$, usually $\alpha \approx 0.001$), the learning rate ($\beta \in [0, 1]$, usually $\beta \approx 1$ for fast learning), and vigilance ($\rho \in [0, 1]$, usually $\rho \approx 0$). On the other hand, fuzzy ARTMAP does not have a mechanism to avoid overfitting noise and applies a hard competitive strategy rather than a soft one, making it more likely to be trapped in local minima [6], [7]. It seems that accuracy is not balanced with transparency, i.e., the knowledge learned by the NN is not converted into meaningful fuzzy if–then rules. Even if a rule-extraction method could be applied, there is no mechanism enforcing compactness, thus, the large number of hidden units produced would undermine the operation. Notably, in RS applications, the number of hidden units (i.e., rules) ranges from over 100 [34] to several hundreds [19].

D. Adaptive-Network-Based Fuzzy Inference Systems (ANFIS)

ANFIS [39] is a five-layer NN architecture that contains both adaptive (having parameters) and fixed nodes (not having parameters). It usually produces TSK-type fuzzy rules, but other types are also possible. ANFIS has been used in pattern recognition for map updating using aerial photographs, with four output classes (road, tree, grass land, and building), giving comparable results to a fuzzy expert system [77]. Benediktsson and Benediktsson [9] also used multiple ANFIS (MANFIS) with four Landsat Multispectral Scanner (MSS) spectral channels plus three topographic input dimensions (elevation, slope, and aspect) to produce ten output land-cover classes. The MANFIS architecture consists of juxtaposed ANFIS systems sharing no modifiable parameters [56]. They conclude that ANFIS is an excellent choice in multisource classification when we are interested in only one class (i.e., binary masks). A major limitation of ANFIS is that it is not suitable for multioutput classification tasks since it only uses one output node. To overcome the limitations, the use of an extended configuration called coactive ANFIS (CANFIS) [54]–[56] is suggested [9]. The power of CANFIS comes from pattern-dependent weights, as opposed to BP-NNs where the weights are used in a global fashion, i.e., the weights are common to all training patterns. CANFIS has an unlimited number of output classes and nonlinear fuzzy rules.

E. VGA and GGAFCM

A series of genetically optimized methods have been applied in order to classify into seven classes a schema that is suitable for pattern recognition rather than for land-use classification (turbid water, pond water, concrete, vegetation, habitation, open space, and roads) [4]. They compared the variable string length GA with chromosome differentiation (VGACD) and the VGA using three-band (G, R, IR) Systeme Pour l'Observation de la Terre (SPOT) multispectral (XS) image data. VGACD and VGA are found to yield comparable results to k-nearest neighbors and MaxLike. In [3], genetic fuzzy clustering is used to overcome fuzzy c-means limitations, namely the requirement of *a priori* knowledge of the number of classes and the frequent trapping in local optima. They use simple (average and difference) kernel (3×3, 5×5, 7×7) context information as they note that it seems natural to take into account not only the spectral properties of the pixels but also their neighbor information. The method is applied on a two-band (G, IR) IRS-1A LISS-II image. It is found that the improvement is negatively related to the amount of averaging, both in terms of kernel size and function used. Furthermore, computing difference values (top down and low left) yields better results than computing average spectral values, although it is hard to evaluate the actual improvement since an accuracy matrix is not provided by the authors. In [61], the fuzzy variable string-length GA (FVGA) [49] is tested using SPOT (G, R, IR), and the results are compared with fuzzy c-means and GGAFCM [36]. The comparison is not made on the basis of ground samples.

A detailed description of GA and VGA and a classification of Landsat-5 for five geologic classes are given in [4]. Elsewhere, IRS and SPOT are used for seven *ad hoc* classes [63] but only visual assessment is performed to compare GA, VGA, k-NN, and MLP. They find that VGA performs better overall and that GAs perform better, in general, compared to the rest. They use principal component analysis as a feature-selection means to reduce the four MSS bands (G, R, NIR, and IR) into two. It should be noted however that feature compression might jeopardize accuracy given that the number of features is already low. VGA is found to be marginally overall more accurate but it contains a class with 44.44% accuracy. The Bayes classifier yields a slightly lower overall accuracy but much more balanced internal distribution. MLP surprisingly fails to classify the easy (since the other four categories are rock related) category vegetation. It is mentioned that the poor results of the MLP may be due to the inappropriate architecture chosen. This might be true for their one-hidden-layer architecture, which has half of the size proposed by the lower bound Kallelopoulos–Wilkinson rule, but is less likely true for their two-hidden-layer architecture, which conforms well within the rule.

F. Neuro-Fuzzy Classification (NEFCLASS)

This is an FNN1 system according to [16], meaning that it has fuzzy weights but real number inputs. It is a neuro-fuzzy system that simultaneously aims for two objectives. First, to classify as accurately as possible, and second to simultaneously build an NN structure that is as transparent as possible. Transparency is sought by keeping the rule set compressed and is achieved via a pruning phase following training [57]. The rule set is directly expressed as fuzzy if–then rules [60] eliminating the need of applying *a posteriori* rule-extraction methods. Training is carried out via a BP-like technique [84] that has been recently extended to perform on both symbolic and numeric data [58].

NEFCLASS has been used to classify remotely sensed imagery into land-use classes [81]. Four-band (R, G, B, IR) IRS-1A LISS-II data are used in conjunction with topographic ancillary information to classify six land-cover types. Ancillary information includes elevation, slope, and aspect derived from a medium-scale conventional map. The rule set produced by NEFCLASS is found to be accurate and transparent, with only seven rules resulting in more than 80% overall accuracy. They find that the reduced number of fuzzy if–then rules in the rule base is a significant aid toward understanding, and thus improving, the classification procedure. The sensitivity of NEFCLASS parameters (MFs, number of fuzzy sets, rule weights, validation method used, etc.) in a spatial, but not RS-related, problem is discussed in [27]. Duch *et al.* [28] extensively compared NEFCLASS with other systems and found that NEFCLASS could further benefit from manual feature selection to yield the same level of accuracy with fewer rules. Elsewhere [31], they comment that the use of context-dependent (different for each rule) linguistic variables in NEFCLASS would probably produce better results.

G. GA-MLP

The synergy of GAs and NNs is, in general, realized in three forms: NNs trained by GAs, trained NNs pruned by GAs, and, finally, NNs trained and pruned by GAs. The first two cases have been exploited with RS data [9], and the conclusion is that results can be superior to standard NNs. Liu *et al.* [47] made two experiments. The first is with a three-band (G, R, IR) SPOT-4 XS using MaxLike, BP-MLP, and GA-MLP. Under the GA-MLP architecture, the NN is trained via a GA and then further trained via BP. GA-MLP is found to be more accurate. The classification schema contains only three categories, which are normally linearly separated (urban, water, and vegetation). The NN architecture used (3:8:3) is the same for BP and GA and falls within the Kanellopoulos–Wilkinson rule. The second experiment involves a China–Brazil ERS (CBERS) image with three bands (G, R, IR) and six output classes. GA and BP NNs yield comparable results.

H. Self-Constructing Neural Fuzzy Inference Network (SONFIN)

SONFIN is a neuro-fuzzy system capable of automatically determining its optimal structure and parameters, thus resulting in a compact network structure [41]. Initially, the internal structure of the NN is blank, and nodes and links are progressively added following a learning process. The SONFIN system has been deployed in the context of RS [46]. A three-band SPOT High Resolution Visible image has been classified into five classes (soil, forest, sea, residential, and industrial). Twelve statistical and 24 spectral contextual features have been added to the spectral information raising the dimensionality of input space to 39. The first two-layered architecture, comprised of an initial SOM followed by a standard BP, yields significant (6%) improvement over a standard k-nearest neighborhood. The second architecture, which consists of the same SOM as the first layer and an SONFIN as the second layer, further improves (by 10% compared to BP) the results in terms of overall accuracy. Another interesting finding is that the SONFIN architecture results into a 20% improvement compared to k-nearest neighborhood in the classification of the two relatively difficult to separate urban classes (residential and industrial).

I. Fuzzy Neural System (FuNe)

FuNe flavors (version I, II, and I-FS) are presented in [35]. The extraction of both positive type ("if") and negative type ("if-not") TSK rules is possible. Three types of rules are generated; simple rules having only a single fuzzy input in the premise, conjunctive (AND) rules, and disjunctive (OR) rules. Optimization includes pruning of rules in the rule set. A four-band IKONOS satellite image has been classified into eight land-use classes in [37]. The neuro-fuzzy classification has been found to be more accurate compared to MaxLike and standard BP-NN methods, although a thorough accuracy assessment in not part of the study. An interesting finding, according to the authors, is that the neuro-fuzzy method performs better for mixed classes such as bare soil and coniferous trees, whereas for homogenous classes, such as "asphalt road," the MaxLike outperforms the rest.

III. Method Comparison

A. Data

The methods are tested using a satellite image of Lefkas island located in the western part of the Hellenic Republic. The satellite image is a seven-band Landsat-7 ETM+ scene acquired on July 2000. In addition, topographic data available under the Shuttle Radar Topography Mission (SRTM) [76] are used. Based on the original digital elevation model (DEM) topographic data, the derivative slope and aspect are constructed. The spatial resolution of the topographic data is approximately 90 m. The classification schema used is that specified by the European Commission program to COoRdinate INformation on the Environment (CORINE) [25]. CORINE level 1 land-use classes are used, namely artificial surfaces (ATF), agricultural areas (AGR), forest and seminatural areas (FOR), wetlands (WET), and water bodies (WAT). Binary coding is used to represent them. In an effort to keep the architecture and rules as compact as possible, we select a subset of input dimensions for subsequent tests. Only Landsat bands 3 (0.63–0.69), 4 (0.76–0.9), and 6 (10.4–12.5) are used, and named RED, NIR, and TIR, accordingly (Fig. 1). More sophisticated feature-selection methods [8] are outside the scope of this paper. In addition, aspect is removed and only elevation (named DEM) and slope (in degrees, named SLOPE) are kept.

We acquired 50 samples per class, resulting in a total of 250 cases. The intended schema is stratified random sampling, although in some cases, especially for wetlands, we were forced to deviate due to the limited area that the class occupies on the ground. We used the CORINE land-cover map as reference data. The available samples are randomly split in half, consisting of 125 samples each, one part being used for training and the other for testing. The two halves are swapped each time a good set of parameters is found for the methods tested in the following sections, to check for sensitivity to the particular samples used.

B. NN Low Bound

Two standard NNs are used as a basis for the comparison of results of different methods. The first is constructed according to the Kanellopoulos–Wilkinson low-bound rule, and the second to the high bound rule. The low-bound NN is described in this section, whereas the high bound is described in the following one.

1) Structure: Given that we have five inputs and five outputs (the numbers are equal by chance), the resulting structure is 5:10:5 (ten hidden-layer nodes). We use a feedforward network trained by BP, as shown in Fig. 2. The training function is batch gradient descent with momentum. The mean squared error (MSE) is used as the performance function. The transfer function in both the hidden and the output layer is log sigmoid.

2) Method and Parameters: Several different parameter configurations are tested. The procedure is to train for a pre-specified error level and then compute the accuracy matrices for

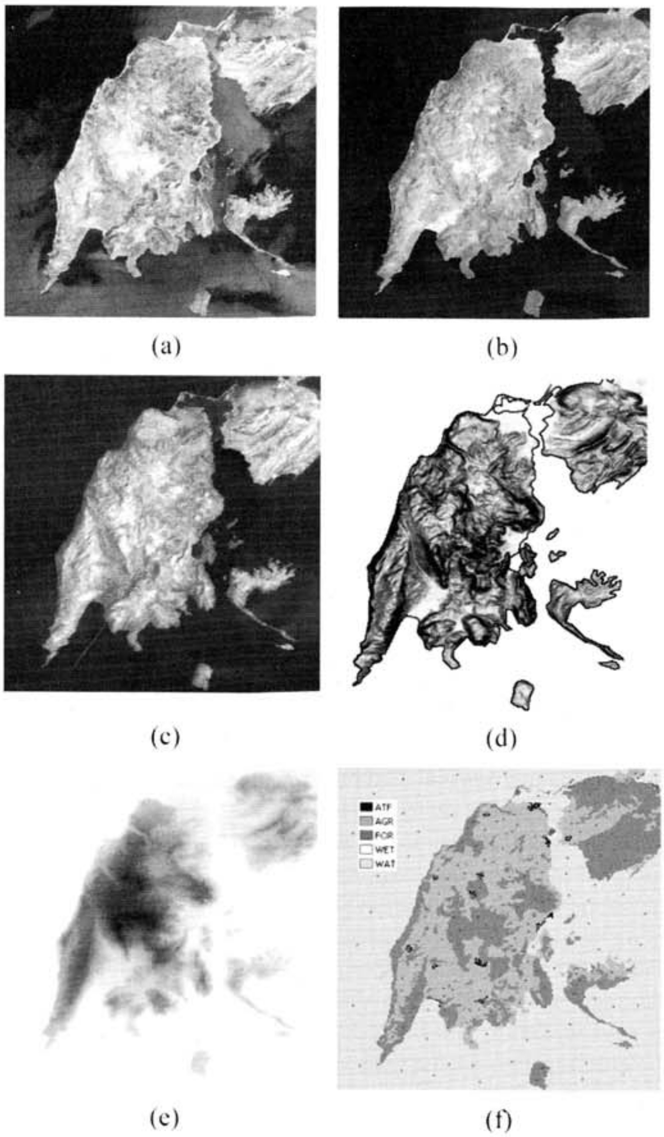

Fig. 1. (a)–(e) Five input parameters used in the classification tests. (f) The samples taken and the CORINE data set. (a) Landsat visible red (RED). (b) Landsat NIR. (c) Landsat TIR. (d) SLOPE, derived from DEM (the darker, the higher the slope is). (e) DEM. (f) Samples and CORINE level 1.

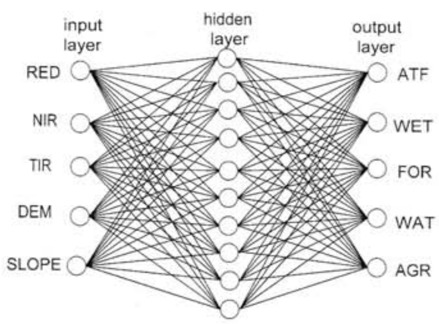

Fig. 2. Architecture of the low-bound 5:10:5 NN.

both training and testing, in order to check the generalization capacity as well as gain understanding of how class separation evolves. It is found that testing-set accuracy initially increases as training accuracy increases. After a certain training accuracy-level threshold, testing accuracy starts degrading. This is the point where training stops. It has proven very useful to look at the training-set accuracy matrix after each stop. This gives an immediate impression on whether the network has matured enough by looking at the internal distribution of error. Sometimes, the overall accuracy hides the fact that samples for certain classes are poorly classified.

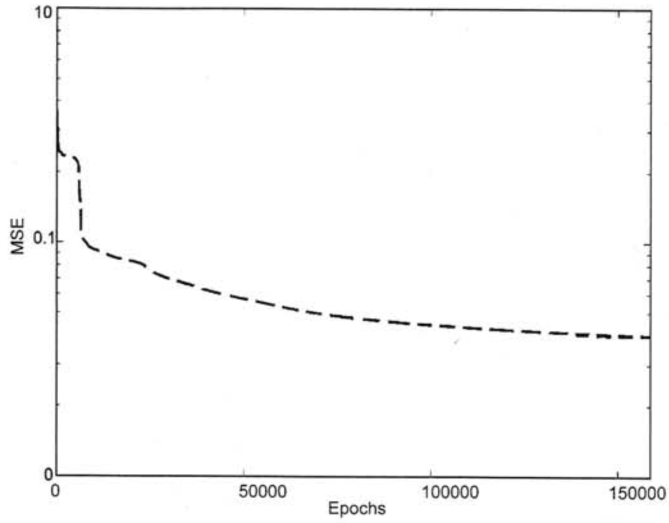

Fig. 3. Training of the 5:10:5 network. The Y axis shows MSE.

According to this training philosophy, the best results are obtained with a learning rate of 0.03, momentum of 0.9, for an error level of 0.04 MSE. Training continues for 158 601 epochs and stops after ten epochs in a row that the error remains stable (Fig. 3). In general, it is found that altering the learning rate and the momentum has an impact on training time but not on accuracy. Regarding training time, it takes approximately 1 min to train for 15 000 epochs.

3) Accuracy Assessment: The results for training and testing are very satisfactory given the limited size of samples per class (25 on average). Accuracy matrices for both training and testing are shown on Table II. Swapping training and testing data sets results in an $\sim 6\%$ reduction in overall testing accuracy.

C. NN High Bound

1) Structure: The structure of the high-bound network in our case, according to the Kanellopoulos–Wilkinson rule, is 5:15:5, as shown in Fig. 4.

2) Method and Parameters: The training strategy is the same as with the low-bound network described in the previous section. Best results are obtained with a learning rate of 0.03, momentum of 0.9, for an error level of 0.04 mse. Training continues for 62 491 epochs and stops after ten epochs in a row that the error remains stable (Fig. 5).

3) Accuracy Assessment: The accuracy matrices for training and testing are shown on Table II. Swapping training and testing data sets results in $\sim 11\%$ reduction in overall testing accuracy.

D. CANFIS

ANFIS, in the light of land-use classification, is very limited due to the fact that it can only have a single output unit.

TABLE II
NEFCLASS RULE BASE FOR ALL CLASSES

RED	NIR	TIR	DEM	SLOPE	OUT
medium	medium	large	medium	very small	ATF
medium	medium	medium	very small	very small	ATF
medium	medium	large	medium	small	ATF
very large	large	large	large	very small	ATF
medium	medium	large	small	small	ATF
medium	small	small	very small	medium	ATF
medium	medium	large	small	very small	ATF
medium	medium	medium	very small	medium	ATF
very small	very small	small	very small	very small	WET
very small	small	small	very small	very small	WET
small	very small	small	very small	very small	WET
very small	very small	small	very small	small	WET
very small	small	medium	very small	small	WET
small	small	medium	very small	very small	WET
medium	small	medium	very small	very small	WET
very small	small	medium	medium	small	FOR
medium	medium	large	very large	small	FOR
very small	small	medium	small	medium	FOR
very small	small	medium	medium	medium	FOR
small	small	medium	very large	small	FOR
medium	medium	large	large	medium	FOR
very small	small	medium	small	small	FOR
small	medium	large	very large	small	FOR
very small	very small	very small	very small	very small	WAT
small	small	large	small	very small	AGR
small	medium	large	very small	very small	AGR
small	small	medium	medium	small	AGR
small	medium	medium	very small	small	AGR
medium	small	large	large	very small	AGR
small	medium	medium	very small	very small	AGR
small	medium	large	very small	small	AGR

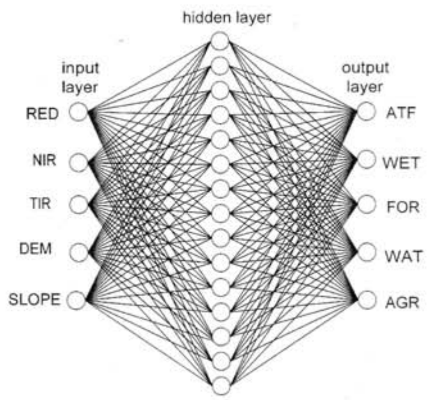

Fig. 4. Architecture of the high-bound 5:15:5 NN.

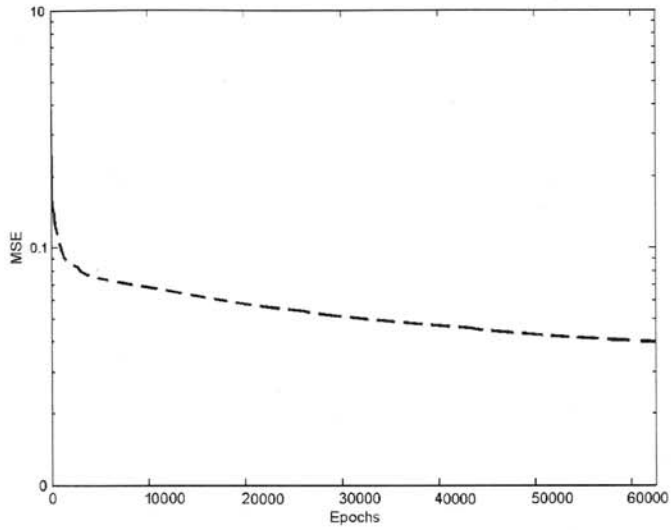

Fig. 5. Training of the 5:15:5 network. The Y axis shows MSE.

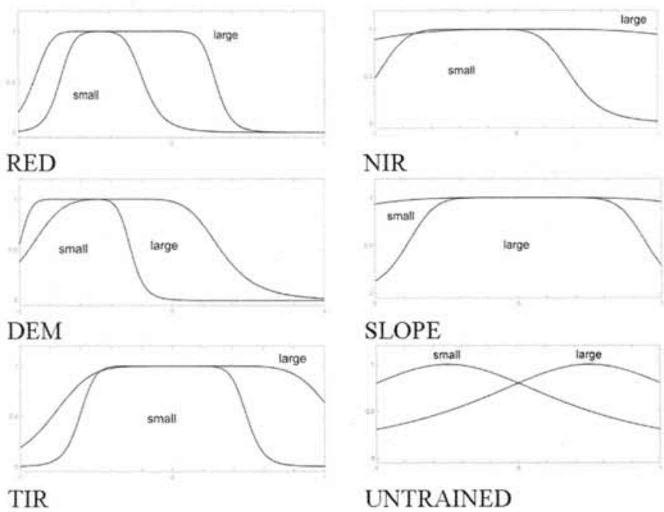

Fig. 6. CANFIS fuzzy sets, before and after training.

Although this could be considered an advantage when constructing masks (e.g., crop masks), it is a major handicap for multiclass schemas. For that reason, we do not experiment with ANFIS or MANFIS, but rather we focus on its multiple-output counterpart CANFIS.

Two bell-shaped fuzzy sets per input are used (Fig. 6). Three or more sets are not feasible in practice because of the significant increase in training time. The stopping criterion is the completion of 10 000 training epochs. To conclude training, 17 min are enough. Momentum is set to 0.7 and learning rate to 1. Weights are batch updated. TSK rules are the outcome of the system. Our particular structure is shown in Fig. 7, and the training progress in Fig. 8. Accuracy matrices for training and testing are given in Table II. The "winner takes all" (WTA) rule is used to produce crisp values. Swapping the data sets results into an $\sim 5\%$ overall performance degradation for the testing data set.

E. Fuzzy ARTMAP

1) Structure and Parameters: The first thing to notice when configuring fuzzy ARTMAP is the very few parameters that have to be tuned. It is found by successive runs that fast learning is optimum (rate $= 1$); reducing the speed of learning leads to performance degradation. The maximum number of categories and epochs has to be specified large enough to allow the algorithm to automatically determine the actual required number. Unlike standard NNs, the number of epochs usually needed in fuzzy ARTMAP is very low. Training time is only a fraction of that required by standard NNs to reach the same level of accuracy. The choice parameter is fixed close to zero (0.000 001).

It has been confirmed in practice that the parameter with the largest impact on accuracy is vigilance, which determines the resolution of the classification processes [34]. Here, we have specified fast learning and fixed the number of maximum categories and epochs to a number higher than that actually required (110 and 100, respectively). We investigated several configurations by gradually increasing the vigilance parameter

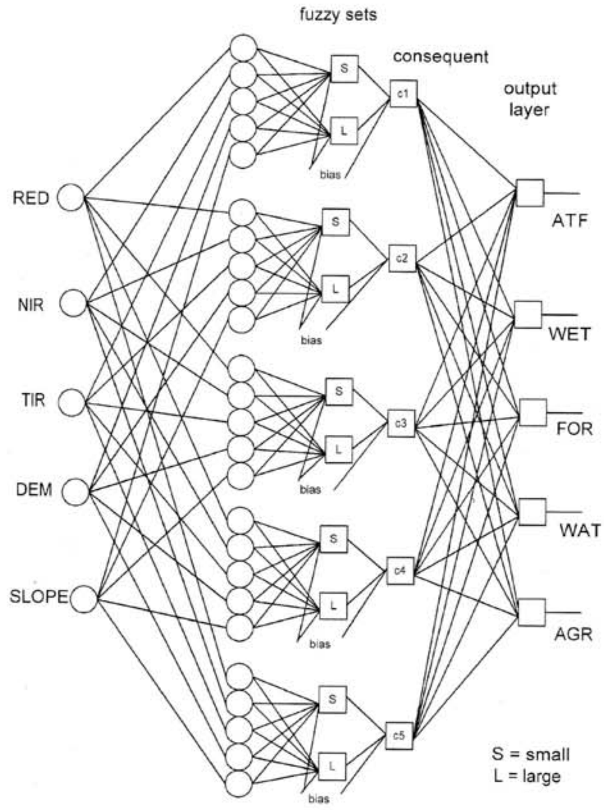

Fig. 7. CANFIS architecture.

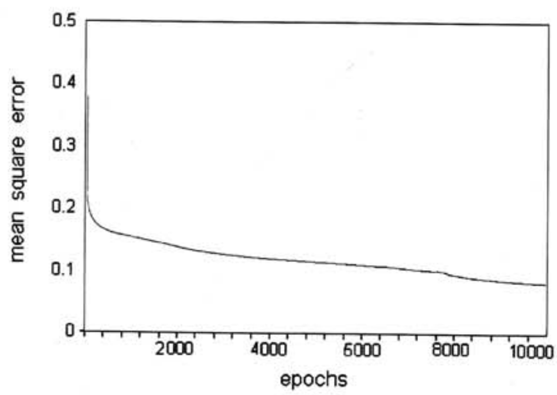

Fig. 8. CANFIS training.

by a small step, as shown in Fig. 9. Best results are observed for vigilance $= 0.9108$. The run with the best results presented here requires 5 epochs and 95 categories for original sets, and 7 epochs and 101 categories after swapping them. Table III summarizes the values of the parameters used here and in other similar cases in the literature.

2) Accuracy Assessment: The training and testing accuracy matrices support the argument that high vigilance values result in pure generalization [34]. All runs exhibit overfitting the training set, which is classified totally correct, but generalization is not as successful. Swapping training and testing data sets results in an $\sim 3\%$ reduction in overall testing accuracy. There are two similar unclassified vectors labeled as wetlands. Accuracy matrices for training and testing are given in Table II.

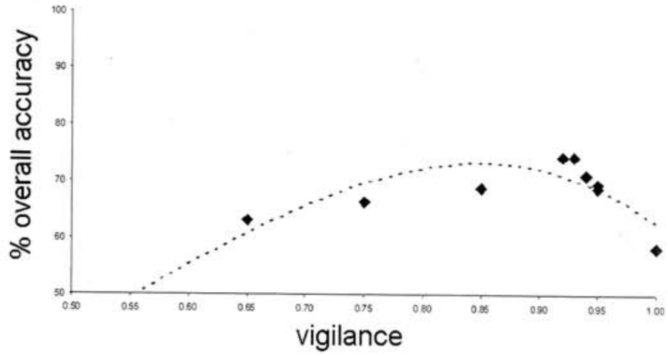

Fig. 9. Fuzzy ARTMAP. Relation of vigilance and percent overall accuracy. Trend line added for clarity.

F. NEFCLASS

Five trapezoidal fuzzy sets are selected per variable (Fig. 10). The aggregation function is maximum and the WTA rule is used to produce a crisp output. The maximum number of rules is set to 50, selecting the best per class. The rules are further reduced to 31 by pruning. The final architecture is shown in Fig. 11. The learning rate is set to 0.5 and fuzzy sets are forced to keep their relative order and always overlap. The maximum number of epochs is intentionally set to a number that is never reached. To maintain interpretability, rule weights are not used. The rules obtained are shown on Table IV. Training time is less than 1 min. We tried manual feature selection to achieve an even more compact rule set but without success. Swapping data sets has no effect on accuracy.

NEFCLASS configurations with loose restrictions on the number of rules or the number of fuzzy sets can yield high accuracy, comparable to that of NNs; however, the cost is reduced transparency. In order to produce a compact rule set, two actions can be taken. The first is to keep the number of fuzzy sets limited, which in our experience means less than five; with more fuzzy sets than that, the rank becomes unclear. The second option is to keep the number of rules limited. The exact number is related to the number of outputs, but more than ten rules per class make the semantics difficult to comprehend. The training strategy can be to let the restrictions loose to begin with and see what the maximum accuracy that the method can reach for the given data set can be. Then, start to limit the number of rules or the number of hedges or both, until a more compact structure is found with an acceptable accuracy level.

G. FuNe

The FuNe I architecture has several adjustable parameters and it is capable of extracting TSK rules. Learning is achieved via BP. Three fuzzy sets per input variable are selected based on the sigmoid MF, as shown on Fig. 12. The architecture is shown on Fig. 13, where the original [35] symbology is kept. Nodes containing $\cup$ and $\cap$ have soft max and soft min operators, accordingly, whereas the rest contain the sum operator. The stopping criterion is a fixed number of training epochs, set to 32 767. Several configurations where tested, mainly by altering learning rate and momentum while keeping everything else fixed. Linear rule nodes are selected because sigmoid rules do

TABLE III
ACCURACY MATRICES, TRAINING LEFT, TESTING RIGHT. PRODUCER'S AND USER'S REFER TO PERCENT ACCURACY

NN LOW

Khat = 0.88	ATF	WET	FOR	WAT	AGR	User's
ATF	24	2	1	0	1	85.7
WET	1	22	0	0	0	95.7
FOR	0	0	18	0	3	85.7
WAT	0	0	0	22	0	100
AGR	1	0	3	0	27	87.1
producer's	92.3	91.7	81.8	100	87.1	90.4

Khat = 0.75	ATF	WET	FOR	WAT	AGR	User's
ATF	15	1	1	0	0	88.2
WET	1	20	0	0	0	95.2
FOR	1	0	23	0	5	79.3
WAT	0	1	0	28	0	96.6
AGR	7	4	4	0	14	48.3
producer's	62.5	76.9	82.1	100	73.7	80.0

NN HIGH

Khat = 0.84	ATF	WET	FOR	WAT	AGR	User's
ATF	22	1	1	0	2	84.6
WET	0	23	0	0	0	100.0
FOR	0	0	17	0	4	81.0
WAT	0	0	0	22	0	100.0
AGR	4	0	4	0	25	75.8
producer's	84.6	95.8	77.3	100	80.6	87.2

Khat = 0.76	ATF	WET	FOR	WAT	AGR	User's
ATF	18	0	1	0	0	94.7
WET	2	22	0	0	0	91.7
FOR	1	1	19	0	5	73.1
WAT	0	0	0	28	0	100.0
AGR	3	3	8	0	14	50.0
producer's	75.0	84.6	67.9	100	73.7	80.8

CANFIS

Khat = 0.86	ATF	WET	FOR	WAT	AGR	User's
ATF	24	1	0	0	1	92.3
WET	1	22	1	0	0	91.7
FOR	1	0	16	0	5	72.7
WAT	0	0	0	22	0	100
AGR	3	0	1	0	27	87.1
producer's	82.8	95.7	88.9	100	81.8	88.8

Khat = 0.77	ATF	WET	FOR	WAT	AGR	User's
ATF	17	0	3	0	0	85
WET	2	24	0	3	0	82.8
FOR	0	1	21	0	4	80.8
WAT	0	0	0	25	0	100
AGR	5	1	4	0	15	60
producer's	70.8	92.3	75	89.3	78.9	81.6

Fuzzy ARTMAP

Khat = 1	ATF	WET	FOR	WAT	AGR	User's
ATF	26	0	0	0	0	100
WET	0	24	0	0	0	100
FOR	0	0	22	0	0	100
WAT	0	0	0	22	0	100
AGR	0	0	0	0	31	100
producer's	100	100	100	100	100	100

Khat = 0.72	ATF	WET	FOR	WAT	AGR	User's
ATF	13	0	1	0	1	86.7
WET	1	22	0	5	0	78.6
FOR	2	2	25	0	4	75.8
WAT	0	0	0	23	0	100.0
AGR	6	2	2	0	14	58.3
producer's	59.1	84.6	89.3	82.1	73.7	77.6

NEFCLASS

Khat = 0.78	ATF	WET	FOR	WAT	AGR	User's
ATF	19	0	3	0	1	82.6
WET	0	24	2	0	0	92.3
FOR	2	0	22	0	4	78.6
WAT	0	1	0	27	0	96.4
AGR	2	0	6	0	11	57.9
producer's	82.6	96.0	66.7	100	68.7	82.4

Khat = 0.66	ATF	WET	FOR	WAT	AGR	User's
ATF	19	3	3	0	2	70.4
WET	8	16	0	0	0	66.7
FOR	2	0	14	0	6	63.6
WAT	0	0	0	22	0	100
AGR	6	2	3	0	20	64.5
producer's	54.3	76.2	70	100	71.4	72.8

FuNe

Khat = 0.60.	ATF	WET	FOR	WAT	AGR	User's
ATF	22	0	0	0	2	92
WET	0	24	1	1	0	92
FOR	0	0	25	0	3	89
WAT	0	1	0	27	0	96
AGR	4	1	4	0	10	53
producer's	85	92	83	96	67	86.4

Khat = 0.60	ATF	WET	FOR	WAT	AGR	User's
ATF	18	2	0	0	6	69
WET	7	17	0	0	0	71
FOR	2	4	13	0	3	59
WAT	0	0	0	22	0	100
AGR	7	3	6	0	15	48
producer's	53	65	68	100	63	68

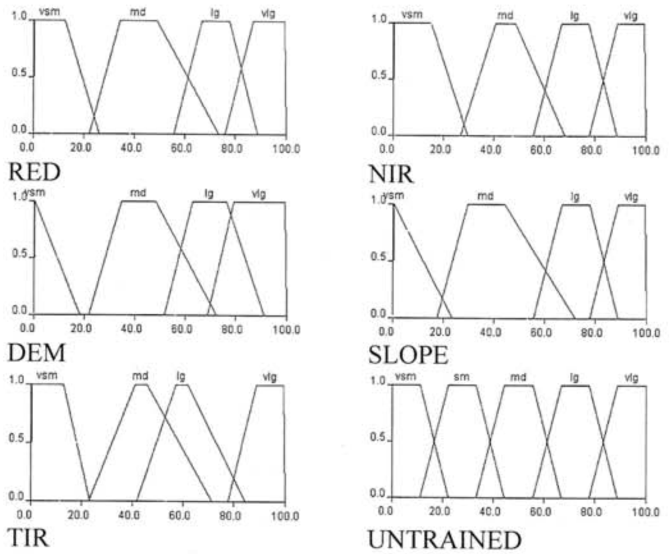

Fig. 10. NEFCLASS fuzzy sets. Relative order of fuzzy stets is preserved. Very small, small, medium, large, and very large are denoted as vsm, sm, md, lg, and vlg, respectively.

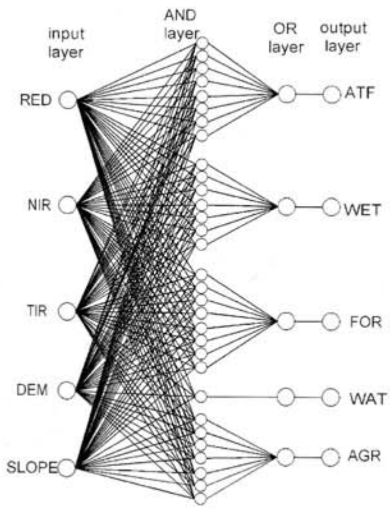

Fig. 11. NEFCLASS architecture. Thirty-one rules are formed after pruning.

TABLE IV
COMMON FUZZY ARTMAP PARAMETER CHOICES IN REMOTE SENSING

	domain	[33]	[26]	[48]	[34]	*
(α)	>0	0	~0.001	~0.01	~0	0.000001
(β)	[0,1]	0.875	~1	1	1	1
(ρ)	[0,1]	0.912	~1	1	0.5	0.9108

* = present work
(α) = Choice parameter, (β) = Learning rate, (ρ) = Vigilance

not yield any performance increase. The number of rules is fixed to 50. The best results are achieved for learning rate set to 0.001 and momentum to 0.6. The tuning of MFs starts after overall error reaches 15%, as shown on Fig. 14. There are no connections from the target layer to bias neurons. The same parameters are used for training and optimizing. The maximum weight value and difference to be neglected is set to 0.01 for both the most rule-relevant node on an input and the second most relevant. A change in sign for the second rule-relevant

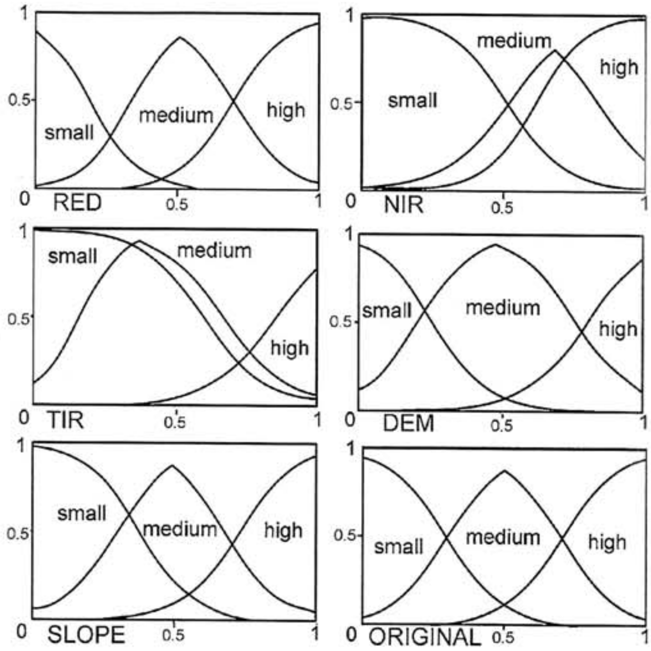

Fig. 12. FuNe fuzzy sets. The Y axis is the degree of membership [0,1]. The X axis is the input range ([0,1]).

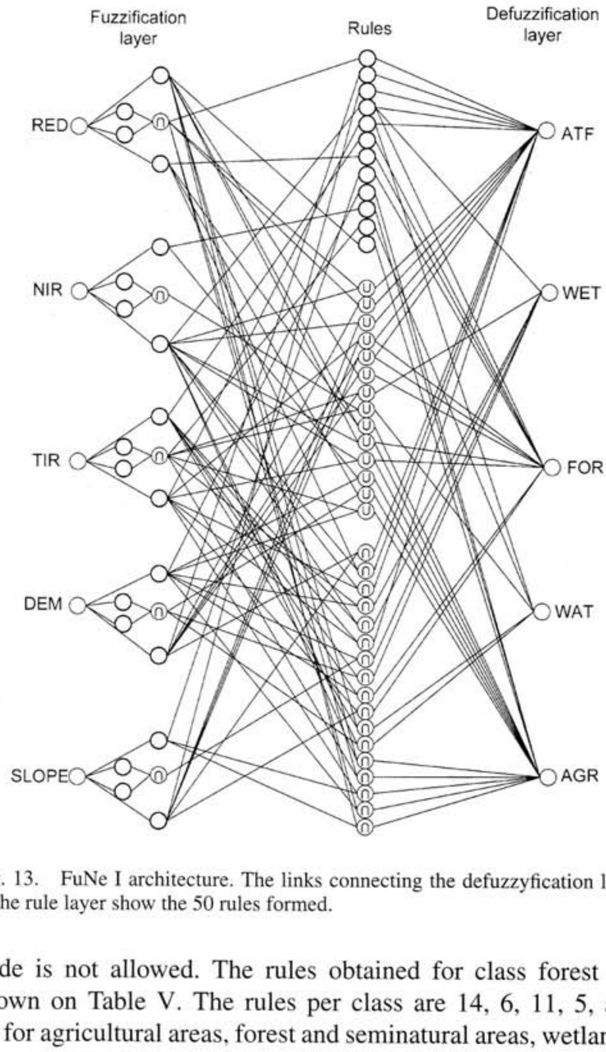

Fig. 13. FuNe I architecture. The links connecting the defuzzyfication layer to the rule layer show the 50 rules formed.

node is not allowed. The rules obtained for class forest are shown on Table V. The rules per class are 14, 6, 11, 5, and 14 for agricultural areas, forest and seminatural areas, wetlands, and water bodies, accordingly. Almost half of them are negative

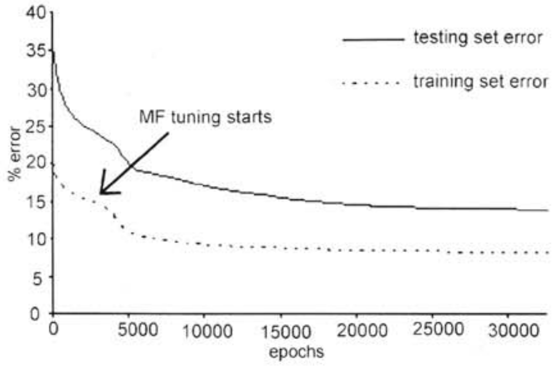

Fig. 14. FuNe percent training error for the pruned structure. After 2800 epochs, the error drops below 15% and tuning of MFs initiates.

TABLE V
FuNe Rule Base for Class Forest

#	WEIGHT	PREMISE	CONCLUSION
1	-2.474	IF TIR IS LOW	THEN NOT FOREST
2	-1.392	IF NIR IS HIGH	THEN NOT FOREST
3	-0.664	IF RED IS HIGH	THEN NOT FOREST
4	-0.623	IF SLOPE IS LOW	THEN NOT FOREST
5	2.322	IF SLOPE IS HIGH AND TIR IS HIGH	THEN FOREST
6	-1.771	IF TIR IS LOW AND DEM IS LOW	THEN NOT FOREST
7	1.917	IF RED IS LOW OR DEM IS HIGH	THEN FOREST
8	1.360	IF RED IS LOW OR NIR IS HIGH	THEN FOREST
9	1.344	IF DEM IS HIGH OR SLOPE IS HIGH	THEN FOREST
10	-0.767	IF NIR IS HIGH OR TIR IS HIGH	THEN NOT FOREST
11	0.161	IF TIR IS HIGH OR SLOPE IS HIGH	THEN FOREST

and half are positive. Interestingly, FuNe needs as much as five rules, two positive and three negative, to classify the easy-category water bodies. The magnitude of the weights shows the firing strength of the rule whereas the sign makes the rules positive or negative. Swapping training and testing data sets results in a 5% loss in overall accuracy.

IV. Comparison of Experimental Results

A comparison of different CI methods has to be done in many dimensions other than accuracy. These other main axes for the comparison can be the three desirable characteristics of fuzzy systems [40].

1) Completeness, evaluating whether all input vectors are classified.
2) Consistency, under which rules with similar premises pose similar consequents and at the same time those rules do not conflict with knowledge of experts.
3) Compactness, is keeping the number of rules as low as possible.

The overall accuracy is obviously important, but because the output is land-cover mapping, the distribution of error is also significant, hence the inclusion of worst class accuracy in Table VI. Accuracies are more appropriately expressed using the k-hat statistic to account for randomness [24]. In our case, the rank of the methods with respect to overall accuracy does not change when taking k-hat into consideration.

CANFIS and NEFCLASS are the most compact configurations used followed by FuNe. For NEFCLASS, we tried to

TABLE VI
Method Comparison for Testing Data Sets

	Low MLP	high MLP	NEF CLASS	Fuzzy ART MAP	CAN FIS	Fune
Khat	0.75	0.76	0.66	0.72	0.77	0.6
overall accuracy	80	80.8	72.8	77.6	81.6	68
worse producer's accuracy[a]	62.5 ATF	67.9 FOR	54.3 ATF	59.1 ATF	70.8 ATF	53 ATF
completeness[b]	100	100	96	98.4	100	100
compactness[c]	n.a.	n.a.	31	n.a.	25	50
consistency[d]	n.a.	n.a.	yes	n.a.	n.t.	no

[a]percentage and worse class are given.
[b]measured as % classified samples (as opposed to unclassified).
[c]measured as the number of rules.
[d]"yes" if the meaning of rules is consistent, "no" if they are contradictory.
n.a = not applicable
n.t. = not tested

have fewer rules, but the overall accuracy drops significantly. CANFIS cannot be more compact because our configuration already contains only two linguistic terms per variable. For FuNe, the accuracy is already too low to opt for a more compact rule set, which normally results to a further decrease in accuracy.

NEFCLASS and fuzzy ARTMAP do not completely classify the samples. Nevertheless, the portion left out is very low, which are two instances for fuzzy ARTMAP and five for NEFCLASS. In the latter case, incompleteness frequently occurs due to rounding precision. Usually, there are two winners, having exactly the same activation in the given precision, after the application of the WTA rule. In the present paper, all of the unclassified instances where manually committed to the wrong category in order to construct the accuracy matrix; fuzzy ARTMAP unclassified instances were simply omitted from the matrices.

NEFCLASS rules are consistent, whereas in FuNe, they are not always consistent. For example, the following rules for agriculture contain the same condition in the premise (if RED is low) but result in contradictory conclusions.

IF RED is low OR RED is high
  THEN NOT Agriculture (weight -1.615)
IF RED is low
  THEN Agriculture (weight 0.565)

There are other instances of contradictions in the FuNe rule set that are not shown here.

NEFCLASS is less sensitive to the specific data used. The high-bound NN and fuzzy ARTMAP are the most data-sensitive configurations. CANFIS is accurate, able to generalize and produce well a balanced internal distribution of error. In general, it might be more appropriate to have three data sets, i.e., training, testing, and validation, consisting of at least 50 samples each. This way, learning would be based on the training data set whereas the stopping criterion could be the overall accuracy of the testing data set. The accuracy reported should then be the one achieved for the validation data set. In other words, the validation data set should be kept completely independent of the training process to estimate more accurately the ability of the method to extrapolate. It is not, however, always feasible to have a large number of samples.

Computational cost is not directly considered in the comparison because in land-cover studies, there is usually no real-time constraint. All the methods tested, with the exception of CANFIS having more than two linguistic terms, converge in reasonable times. The increased processing power of current-generation personal computers has progressively made training time a less important factor in selecting the most suitable method.

V. Conclusion

To the best of our knowledge, this is the first systematic approach to evaluate several CI methods in RS. Most of the existing CI applications in the field, which in any case are not numerous, are focused on pattern recognition rather than land-cover classification, where the context is clearly different. The majority of the existing papers consist of preliminary work without thorough assessment of results. It is interesting to note that there are hardly any papers to show extensive comparison results between two different CI methods, except in the case of comparisons with NNs.

Another interesting aspect of the current status of CI applications in RS is that almost all the applications do not go beyond accuracy to investigate transparency. Fuzzy rules are rarely the outcome of the process, which contradicts the perception that what is expected from CI methods is increased transparency rather than increased accuracy. As our results indicate, if we are only interested in accuracy, there is hardly any need to look further than NNs alone, even when using heuristics to determine their structure and parameters.

The quest for finding optimum CI configurations requires a quite different approach compared to that employed in standard NNs. The objective of training is shifted towards obtaining minimal topology (fewer rules) rather than trying varying topologies in order to achieve optimum performance. The rationale of including FL in the algorithms is to be able to capture the semantics of the classification. If the number of rules per class grows too high, then it is not possible to gain a practical understanding of how the classification works. What governs the final selection of topology is constantly reducing the number of rules until accuracy drops below the level set as acceptable. The classification is not successful unless an acceptable level of accuracy is achieved concurrently with achieving a reasonably low amount of rules per class. If that is not possible, we do not have the option to expand the topology, as we would do with standard NNs, because then the transparency benefit of using FL in the process is lost.

The impact of experimenting with the training termination condition (from minimal to optimal training) is far less critical compared to standard NNs. For example, NEFCLASS accepts a maximum epoch stopping criterion, but training lasts, in general, just a few epochs so that it is not practical to use it against overfitting. Nevertheless, the generalization capacity is usually preserved by limiting the number of rules in the rule base rather than limiting the duration of training. In other words, it is the compact structure, necessary for NNs to be transparent, that prevents overfitting the training data.

Although there is no explicit provision for handling the noise naturally present in any spatial data, the results show that most of the methods perform robustly in the presence of small amounts of noise. The main sources of noise in this paper have been measurement accuracy in collecting the samples, image misregistration, and most particularly, the definition of classes in the land-cover data set used for annotating the output vectors. Although their existence is indisputable, we believe that these sources of error have been kept to a minimum here.

In addition, CI is found to be a suitable paradigm for multisource classification, which is becoming increasingly important as the availability of ancillary data sets, such as SRTM topographic data, continues to grow. It seems unwise to neglect the vast additional information that is now available in any RS classification method. All the methods have been tested here with input vectors obtained by fusing spectral and topographic information. Consequently, the basic principle of data fusion is in place, which could be also applied to handle multisensor RS imagery in future studies.

The results presented here are by no means exhaustive but rather a first effort to bring several methods together and test them in order to highlight some of their weaknesses and strengths. One of the possibly most discouraging aspects of CI for the newcomer in the field is the very high number of different methods and variants. We have tried to select the most important among them and present them in the limited space of this paper. These results suggest that two neuro-fuzzy techniques, namely NEFCLASS and CANFIS, should receive more attention.

Acknowledgment

The authors would like to gratefully acknowledge the aid of S. Kay and the three anonymous reviewers in improving this manuscript. The intellectual property rights of the satellite image used belong to IMAGE2000 of JRC, based on Landsat-7 ETM+ ESA distributed by Eurimage; orthocorrection EU15 Metria, orthocorrection other countries GISAT; mosaic production GISAT. CORINE is "EEA, Copenhagen, 2000" downloaded from http://dataservice.eea.eu.int/dataservice/metadetails.asp?id=759.

References

[1] A. Abuelgasim, W. Ross, S. Gopal, and C. Woodcock, "Change detection using adaptive fuzzy neural networks: Environmental damage assessment after the gulf war," *Remote Sens. Environ.*, vol. 70, no. 2, pp. 208–223, Nov. 1999.

[2] G. Amici, F. Dell'Acqua, P. Gamba, and G. Pulina, "A comparison of fuzzy and neuro-fuzzy data fusion for flooded area mapping using SAR images," *Int. J. Remote Sens.*, vol. 25, no. 20, pp. 4425–4430, Oct. 2004.

[3] S. Bandyopadhyay, "Satellite image classification using genetically guided fuzzy clustering with spatial information," *Int. J. Remote Sens.*, vol. 26, no. 3, pp. 579–593, Feb. 2005.

[4] S. Bandyopadhyay and S. Pal, "Pixel classification using variable string genetic algorithms with chromosome differentiation," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 303–308, Feb. 2001.

[5] A. Baraldi, E. Binaghi, P. Blonda, P. A. Brivio, and A. Rampini, "Comparison of the multilayer perceptron with neuro-fuzzy techniques in the estimation of cover class mixture in remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 5, pp. 994–1005, May 2001.

[6] A. Baraldi and P. Blonda, "A survey of fuzzy neural networks for pattern recognition—Part I," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 29, no. 6, pp. 778–785, Dec. 1999.
[7] ——, "A survey of fuzzy neural networks for pattern recognition—Part II," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 29, no. 6, pp. 786–801, Dec. 1999.
[8] J. A. Benediktsson and J. R. Sveinsson, "Feature extraction for multisource data classification with artificial neural networks," *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 727–740, Mar. 1997.
[9] J. A. Benediktsson and H. Benediktsson, "Neuro-fuzzy and soft computing in classification of remote sensing data," in *Proc. SPIE, EUROPTO Conf. on Image and Signal Processing for Remote Sensing V*, vol. 3871, Florence, Italy, 1999, pp. 176–187.
[10] J. A. Benediktsson, P. H. Swain, and O. K. Ersoy, "Neural network approaches versus statistical methods in classification of multisource remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 4, pp. 540–552, Jul. 1990.
[11] J. Benitez, J. Castro, and I. Requena, "Are artificial neural networks black boxes?," *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 1156–1164, Sep. 1997.
[12] J. Bezdek, T. Reichherzer, G. Lim, and Y. Attikiouzel, "Multiple-prototype classifier design," *IEEE Trans. Syst., Man, Cybern C, Appl. Rev.*, vol. 28, no. 1, pp. 67–79, Feb. 1998.
[13] P. Blonda, A. Bennardo, G. Satalino, and G. Pasquariello, "Fuzzy logic and neural techniques integration: An application to remotely sensed data," *Pattern Recognit. Lett.*, vol. 17, no. 13, pp. 1343–1348, Nov. 1996.
[14] G. Bortolan, "An architecture of fuzzy neural networks for linguistic processing," *Fuzzy Sets Syst.*, vol. 100, no. 1–3, pp. 197–215, Nov. 1998.
[15] G. Bortolan and W. Pedrycz, "Linguistic neurocomputing: The design of neural networks in the framework of fuzzy sets," *Fuzzy Sets Syst.*, vol. 128, no. 3, pp. 389–412, Jun. 2002.
[16] J. Buckley and Y. Hayashi, "Fuzzy neural networks: A survey," *Fuzzy Sets Syst.*, vol. 66, no. 1, pp. 1–13, Aug. 1994.
[17] G. Carpenter, M. Gjaja, S. Gopal, and C. Woodcock, "ART neural networks for remote sensing: Vegetation classification from Landsat TM and terrain data," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 2, pp. 308–325, Mar. 1997.
[18] G. Carpenter, B. Milenova, and B. Noeske, "Distributed ARTMAP: A neural network for fast distributed supervised learning," *Neural Netw.*, vol. 11, no. 5, pp. 793–813, Jul. 1998.
[19] G. Carpenter, S. Gopal, S. Macomber, S. Martens, C. Woodcock, and J. Franklin, "A neural network method for efficient vegetation mapping," *Remote Sens. Environ.*, vol. 70, no. 3, pp. 326–338, Dec. 1999.
[20] G. Carpender, S. Grossberg, and J. Reynolds, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network," *Neural Netw.*, vol. 4, no. 5, pp. 565–588, 1991.
[21] G. Carpenter, S. Grossberg, N. Markuzon, J. Reynolds, and D. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Trans. Neural Netw.*, vol. 3, no. 5, pp. 698–713, Sep. 1992.
[22] I.-F. Chung, C.-J. Lin, and C.-T. Lin, "A GA-based fuzzy adaptive learning control network," *Fuzzy Sets Syst.*, vol. 112, no. 1, pp. 65–84, May 2000.
[23] A. Comber, P. Fisher, and R. Wadsworth, "What is land cover?," *Environ. Plann., B Plann. Des.*, vol. 32, no. 2, pp. 199–209, Mar. 2005.
[24] R. Congalton, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, LA: Lewis, 1998.
[25] CORINE. [Online]. Available: http://image2000.jrc.it/
[26] E. Costamagna, P. Gamba, P. Lombardo, and G. Chinino, "Statistical analysis and neuro-fuzzy classification of polarimetric images of urban areas," in *Proc. ERS/ENVISAT*, Goteborg, Sweden, Oct. 2000.
[27] B. Dixon, "Applicability of neuro-fuzzy techniques in predicting ground-water vulnerability: A GIS-based sensitivity analysis," *J. Spat. Hydrol.*, vol. 4, no. 2, pp. 1–38, 2004.
[28] W. Duch, R. Setiono, and J. Zurada, "Computational intelligence methods for rule-based data understanding," *Proc. IEEE*, vol. 92, no. 5, pp. 771–805, May 2004.
[29] W. Duch, R. Adamczak, and K. Grabczewski, "Extraction of logical rules from backpropagation networks," *Neural Process. Lett.*, vol. 7, no. 3, pp. 1–9, Jun. 1998.
[30] W. Duch, R. Adamczak, K. Grabczewski, and N. Jankowski, "Neural methods of knowledge extraction," *Control Cybern.*, vol. 29, no. 4, pp. 997–1018, 2000.
[31] W. Duch, R. Adamczak, and K. Grabczewski, "A new methodology of extraction, optimization and application of crisp and fuzzy logical rules," *IEEE Trans. Neural Netw.*, vol. 11, no. 2, pp. 1–31, Mar. 2000.
[32] W. Duch and J. Mandziuk, "Quo Vadis computational intelligence," in *Machine Intelligence. Quo Vadis?* vol. 21, *Advances in Fuzzy Systems—Applications and Theory*, P. Sinčák, J. Vaščák, and K. Hirota, Eds. Singapore: World Scientific, 2004, pp. 3–28.
[33] P. Gamba and F. Dell'Acqua, "Increased accuracy multiband urban classification using neuro-fuzzy classifier," *Int. J. Remote Sens.*, vol. 24, no. 4, pp. 827–834, Feb. 2003.
[34] S. Gopal, C. Woodcock, and A. Stahler, "Fuzzy neural network classification of global land cover from 1° AVHRR data set," *Remote Sens. Environ.*, vol. 67, no. 2, pp. 230–243, Feb. 1999.
[35] S. Halgamuge and M. Glesner, "Neural networks in designing fuzzy systems for real world applications," *Fuzzy Sets Syst.*, vol. 65, no. 1, pp. 1–12, Jul. 1994.
[36] L. Hall, I. Ozyurt, and J. Bezdek, "Clustering with a genetically optimized approach," *IEEE Trans. Evol. Comput.*, vol. 3, no. 2, pp. 103–112, Jul. 1999.
[37] J. Han, S. Lee, K. Chi, and K. Ryu, "Comparison of neuro-fuzzy, neural network and maximum likelihood classifiers for land cover classification using IKONOS multispectral data," in *Proc. IGARSS*, Toronto, ON, Canada, 2002, vol. 6, pp. 3471–3473.
[38] J. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor: Univ. of Michigan Press, 1975, p. 211.
[39] J.-S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference systems," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 3, pp. 665–685, May/Jun. 1993.
[40] Y. Jin, W. von Seelen, and B. Sendhoff, "On generating FC^3 fuzzy rule systems from data using evolution strategies," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 29, no. 6, pp. 829–845, Dec. 1999.
[41] C. Juang and C. Lin, "An on-line self-constructing neural fuzzy inference network and its applications," *IEEE Trans. Fuzzy Syst.*, vol. 6, no. 1, pp. 12–32, Feb. 1998.
[42] I. Kanellopoulos and G. Wilkinson, "Strategies and best practice for neural network image classification," *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 711–725, Mar. 1997.
[43] T. Kohonen, *Self-Organizing and Associate Memory*, 3rd ed. New York: Springer-Verlag, 1989.
[44] P. Korning, "Training of neural networks by means of genetic algorithms working on very long chromosomes," *Int. J. Neural Syst.*, vol. 6, no. 3, pp. 299–316, Sep. 1995.
[45] S. C. Lee and E. T. Lee, "Fuzzy neural networks," *Math. Biosci.*, vol. 23, no. 1/2, pp. 151–177, Feb. 1975.
[46] C.-T. Lin, Y.-C. Lee, and H.-C. Pu, "Satellite sensor image classification using cascaded architecture of neural fuzzy network," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 1033–1043, Mar. 2000.
[47] Z. Liu, A. Liu, C. Wang, and Z. Niu, "Evolving neural network using real coded genetic algorithm (GA) for multispectral image classification," *Future Gener. Comput. Syst.*, vol. 20, no. 7, pp. 1119–1129, Oct. 2004.
[48] B. Mannan, J. Roy, and A. K. Ray, "Fuzzy ARTMAP supervised classification of multi-spectral remotely-sensed images," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 767–774, Mar. 1998.
[49] U. Maulik and S. Bandyopadhyay, "Fuzzy partitioning using a real-coded variable length genetic algorithm for pixel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 5, pp. 1075–1081, May 2003.
[50] M. L. Minsky and S. Papert, *Perceptrons: An Introduction to Computational Geometry*. Cambridge: MIT Press, 1969.
[51] S. Mitra and S. K. Pal, "Fuzzy sets in pattern recognition and machine intelligence," *Fuzzy Sets Syst.*, vol. 156, no. 3, pp. 381–386, Dec. 2005.
[52] S. Mitra, R. K. De, and S. K. Pal, "Knowledge-based fuzzy MLP for classification and rule generation," *IEEE Trans. Neural Netw.*, vol. 8, no. 6, pp. 1338–1350, Nov. 1997.
[53] S. Mitra, S. K. Pal, and P. Mitra, "Data mining in soft computing framework: A survey," *IEEE Trans. Neural Netw.*, vol. 13, no. 1, pp. 3–14, Jan. 2002.
[54] E. Mizutani, "Coactive neuro-fuzzy modeling: Toward generalized ANFIS," in *Neurofuzzy and Soft Computing*, J.-R. Jang, Ed. Upper Saddle River, NJ: Prentice-Hall, 1997, ch. 13.
[55] ——, "Soft computing for color recipe prediction," in *Neurofuzzy and Soft Computing*, J.-R. Jang, Ed. Upper Saddle River, NJ: Prentice-Hall, 1997, ch. 22.
[56] E. Mizutani and J.-S. R. Jang, "Coactive neural fuzzy modeling," in *Proc. IEEE Int. Conf. Neural Netw.*, 1995, vol. 2, pp. 760–765.
[57] D. Nauck, "Fuzzy data analysis with NEFCLASS," *Int. J. Approx. Reason.*, vol. 32, no. 2/3, pp. 103–130, Feb. 2003.
[58] ——, "Neuro-fuzzy learning with symbol and numeric data," *Soft Comput.*, vol. 8, no. 6, pp. 383–396, 2004.

[59] D. Nauck and R. Kruse, "Neuro-fuzzy systems for function approximation," *Fuzzy Sets Syst.*, vol. 101, no. 2, pp. 261–271, Jan. 1999.

[60] ——, "A neuro-fuzzy method to learn fuzzy classification rules from data," *Fuzzy Sets Syst.*, vol. 89, no. 3, pp. 277–288, Aug. 1997.

[61] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik, "A study of some fuzzy cluster validity indices, genetic clustering and application to pixel classification," *Fuzzy Sets Syst.*, vol. 155, no. 2, pp. 191–214, Oct. 2005.

[62] S. Pal and S. Mitra, "Multi-layer perceptron, fuzzy sets, and classification," *IEEE Trans. Neural Netw.*, vol. 3, no. 5, pp. 683–697, Sep. 1992.

[63] S. K. Pal, S. Bandyopadhyay, and C. A. Murthy, "Genetic classifiers for remotely sensed images: Comparison with standard methods," *Int. J. Remote Sens.*, vol. 22, no. 13, pp. 2545–2569, Sep. 2001.

[64] N. R. Pal, A. Laha, and J. Das, "Designing fuzzy rule based classifier using self-organizing feature map for analysis of multispectral satellite images," *Int. J. Remote Sens.*, vol. 26, no. 10, pp. 2219–2240, May 2005.

[65] M. Pax-Lenney, C. Woodcock, S. Macomber, S. Gopal, and C. Song, "Forest mapping with a generalized classifier and Landsat TM data," *Remote Sens. Environ.*, vol. 77, no. 3, pp. 241–250, Sep. 2001.

[66] W. Pedrycz, "Fuzzy neural networks and neurocomputations," *Fuzzy Sets Syst.*, vol. 56, no. 1, pp. 1–28, May 1993.

[67] W. Pedrycz and A. Vasilakos, "Linguistic models and linguistic modeling," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 29, no. 6, pp. 745–757, Dec. 1999.

[68] W. Pedrycz and G. Vulkovich, "Granular neural networks," *Neurocomputing*, vol. 36, no. 1–4, pp. 205–224, Feb. 2001.

[69] W. Pedrycz and M. Reformat, "Genetically optimized logic models," *Fuzzy Sets Syst.*, vol. 150, no. 2, pp. 351–371, Mar. 2005.

[70] F. Qiu and J. R. Jensen, "Opening the black box of neural networks for remote sensing image classification," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1749–1768, May 2004.

[71] D. E. Rumelhart, J. McClelland, and PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, Cambridge, MA: MIT, 1986.

[72] M. Russo, "Distributed fuzzy learning using the MULTISOFT machine," *IEEE Trans. Neural Netw.*, vol. 12, no. 3, pp. 475–484, May 2001.

[73] ——, "FuGeNeSys—A fuzzy genetic neural system for fuzzy modeling," *IEEE Trans. Fuzzy Syst.*, vol. 6, no. 3, pp. 373–388, Aug. 1998.

[74] ——, "Genetic fuzzy learning," *IEEE Trans. Evol. Comput.*, vol. 4, no. 3, pp. 259–273, Sep. 2000.

[75] A. F. Shapiro, "The merging of neural networks, fuzzy logic, and genetic algorithms," *Insur., Math. Econ.*, vol. 31, no. 1, pp. 115–131, Aug. 2002.

[76] Shuttle Radar Topography Mission (SRTM), (2000). [Online]. Available: http://www2.jpl.nasa.gov/srtm/ 2000

[77] C. H. Teng and D. Fairbairn, "Comparing expert systems and neural fuzzy systems for object recognition in map dataset revision," *Int. J. Remote Sens.*, vol. 23, no. 3, pp. 555–567, Feb. 2002.

[78] E. C. Tsao, J. C. Bezdek, and N. R. Pal, "Fuzzy Kohonel clustering network," *Pattern Recognit.*, vol. 27, no. 5, pp. 757–764, 1994.

[79] B. C. K. Tso and P. M. Mather, "Classification of multisource remote sensing imagery using a genetic algorithm and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1255–1260, May 1999.

[80] H. Tsukimoto, "Extracting rules from trained neural networks," *IEEE Trans. Neural Netw.*, vol. 11, no. 2, pp. 377–389, Mar. 2000.

[81] A. Vasilakos and D. Stathakis, "Granular neural networks for land use classification," *Soft Comput.*, vol. 9, no. 5, pp. 332–340, 2005.

[82] P. Vuorimaa, "Fuzzy self-organizing map," *Fuzzy Sets Syst.*, vol. 66, no. 2, pp. 223–231, Sep. 1994.

[83] F. Wang, "Fuzzy supervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 2, pp. 194–201, Mar. 1990.

[84] L.-X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 6, pp. 1414–1427, Nov./Dec. 1992.

[85] P. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences," Ph.D. dissertation, Appl. Math. Dept., Harvard Univ., Cambridge, MA, 1974, unpublished.

[86] G. Wilkinson, "Results and implications of a study of fifteen years of satellite image classification experiments," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 433–440, Mar. 2005.

[87] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.

[88] ——, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Syst.*, vol. 90, no. 2, pp. 111–127, Sep. 1997.

[89] ——, "From computing with numbers to computing with words—From manipulation of measurements to manipulation of perceptions," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 46, no. 1, pp. 105–119, Jan. 1999.

[90] M. Michel, *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT, 1996.

Demetris Stathakis received the B.S. degree in planning and regional development engineering from the University of Thessaly, Volos, Hellenic Republic, in 1994, the M.Sc. degree in geographic and geodetic information systems from University College London, London, U.K., in 1995, and the Ph.D. degree from the University of Thessaly, in 2003.

Currently, he is with the Joint Research Centre of the European Commission as a Post-Doc Researcher for monitoring agriculture with remote sensing (RS) applied in food security. His research interests include geographical information systems as well as land-use and land-cover classification using remotely sensed images. He is currently focusing on the application of computational intelligence techniques, namely neural networks, fuzzy sets, and genetic algorithms in RS classification tasks.

Athanassios Vasilakos was born in Pella, Greece. He received the B.S. degree in electrical and computer engineering from the University of Thrace, Xanthi, Greece, in 1983, the M.S. degree in computer engineering from the University of Massachusetts, Amherst, in 1986, and the Ph.D. degree in computer engineering from the University of Patras, Patras, Greece, in 1988.

From 1988 to 1991, he was with Computer Technology Institute, University of Patras. From 1991 to 1996, he was a Professor at the Hellenic Air Force Academy. From 1996 to 2000, he was with the Computer Science Institute, FORTH, Greece. From 2000 to 2005, he was a Professor at the University of Thessaly, Volos, Greece. Since then, he has been a Professor at the Department of Computer and Telecommunications Engineering, University of Western Macedonia, Greece. He has held several visiting Professorship positions. He is the coauthor (with W. Pedrycz) of the books *Computational Intelligence in Telecommunications Networks* (CRC Press, 2001); *Ambient Intelligence, Wireless Networking, Ubiquitous Computing* (Artech House, 2006), and the coauthor of (with M. Parashar, S. Karnouskos, W. Pedrycz) *Autonomic Communications* (Artech House, to appear), *Technology and Theater* (in Greek). He has published more than 100 articles in top international journals and conferences in the area of telecommunications networks, computational intelligence, computers in education and technology and arts. He is a member of the editorial board of several international journals such as *Computer Communications* and *ACM Applied Computing*. He is a reviewer for *Soft Computing, Information Sciences*, the *Journal of Applied Intelligence*, the *International Journal on Computational Intelligence Research*, and the *International Journal on Cognitive Informatics and Intelligence*. He has been a Guest Editor for *Soft Computing* and the *Journal on Interactive Learning Research*.

Dr. Vasilakos was the Chairman of the Telecommunications Committee of the ERUDIT, which is the European network of excellence in fuzzy logic and its applications. He is also the Chairman of the Telecommunications Committee of the IEEE Computational Intelligence Society. He is a reviewer of the IEEE COMMUNICATIONS MAGAZINE and has been a Guest Editor of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS and the IEEE TRANSACTIONS ON NEURAL NETWORKS. He is a member of the ACM.