Global Elevation Ancillary Data for Land-use Classification Using Granular Neural Networks

Demetris Stathakis and Ioannis Kanellopoulos

Abstract

The development of digital global databases containing data such as elevation and soil can greatly simplify and aid in the classification of remotely sensed data to create land-use classes. An efficient method that can simultaneously handle diverse input dimensions can be formed by merging fuzzy logic and neural networks. The so-called granular or fuzzy neural networks are able not only to achieve high classification levels, but at the same time produce compressed and transparent neural network skeletons. Compression results in reduced training times, while transparency is an aid for interpreting the structure of the neural network by translating it into meaningful rules and vice versa. The purpose of this paper is to provide some initial guidelines for the construction of granular neural networks in the remote sensing context, while using global elevation ancillary data within the classification process.

Introduction

The application of neural networks (NNs) in the classification of satellite imagery is one of the benefits in the quest for establishing more efficient methods. There are some clear benefits over statistical classifiers; most of them stem from the fact that NNs make no assumptions about the underlying distribution of the data. The incorporation of ancillary data in the classification becomes therefore straightforward. Ancillary data can be treated as if there were additional spectral bands (i.e., "stacked – vector" approach) without any conceptual incompatibility issues. Overall, the use of ancillary data is sought for two reasons. First, it provides additional information, usually highly uncorrelated to the spectral information at hand, which can be used to either classify more accurately or separate categories that are otherwise difficult to distinguish. Second, the advances in geographical information systems (GIS) have resulted in the development of many databases, some readily available over the Internet, occasionally at no cost. Geospatial technologies are now well advanced and permit the seamless integration of ancillary data held in a GIS database with any raster satellite data. In a way, we can look at ancillary data as a plug-in to any classification problem. It is information that significantly increases the dimensionality of our input, and it is readily available. Early work on the classification of multi-source remotely sensed and geographic data using NNs is presented in Benediktsson *et al.* (1990; 1997).

Some shortcomings of NNs have been noted in recent years. A number of parameters have to be set by trial and error, most likely by using an educated guess based on experience rather than formal rules. The refinement of those parameters clearly calls for guidance to establish a best practice in the classification problem using NNs (Kanellopoulos *et al.*, 1997). Neural networks can be very accurate, but in general, the more accurate the output, the more complex and difficult it is to explain the plethora of weights and connections that form the NN skeleton. This characteristic of NNs is often referred to in the literature as the "black box" syndrome (Benitez *et al.*, 1997). Clearly, if we are to improve and more importantly trust them, we need to understand the processes taking place inside the NNs.

Research to merge fuzzy set theory (Zadeh, 1965) with standard NNs promised a way of creating transparent nets that we are able to interpret. The term Fuzzy Neural Networks (FNN) has been coined a long time ago (Lee *et al.*, 1975). The same basic formation is also referred to as Granular Neural Networks (GNN) (Pedrydz *et al.*, 2001; Bortolan, 1998). GNNs are neural networks in which the processing is done on information granules rather than individual values and fuzzy set theory is deployed to form the granules (Zadeh, 1997). There are several advantages attributed to GNNs. The first comes from the fact that training granules is faster than training original values because the information is reduced (higher level of abstraction). Furthermore, processing is done on a conceptual rather than a numerical level, which opens the way for linguistic processing (Pedrycz *et al.*, 1999) and computing with words (Zadeh, 1999). Finally, information granulation permits building transparent networks. The neural network skeleton is directly converted into a fuzzy rule set (Zadeh, 1997). There are clear guidelines for converting the rule set to neural network skeleton components and *vice versa*. Nevertheless, many other parameters remain to be set. Some of these parameters are common to standard NNs (e.g., learning rate) whereas others are new (e.g., fuzzy shape function).

Demetris Stathakis is with the Institute for the Protection and the Security of the Citizen/MARS/FOOD, European Commission – DG Joint Research Centre, Via Enrico Fermi1 1, Ispra (VA), Italy (demetris.stathakis@jrc.it).

Ioannis Kanellopoulos is with the Institute for Environment and Sustainability, European Commission – DG Joint Research Centre, Via Enrico Fermi1 1, Ispra (VA), Italy.

Photogrammetric Engineering & Remote Sensing
Vol. 74, No. 1, January 2008, pp. 000–000.

0099-1112/08/7401–0000/$3.00/0

The following sections are the application of GNNs in the remote sensing context. The general architecture of the GNN is discussed and established followed by a case study on the application of the GNN. Results including accuracy assessment are given followed by a discussion focusing on some guidelines for the selection of GNN parameters. Finally, conclusions are drawn in the last section with respect to the main purpose of this paper, which is to provide guidelines for the application of GNN in satellite image classification using ancillary information for land-use.

GNN Architecture in the Context of Remote Sensing

The GNN structure in the remote sensing classification context is composed of four layers. Let us refer to them as "fuzzyfication layer," AND layer, OR layer, and de-fuzzyfication layer so that the names correspond to the actual operation (Figure 1). Fuzzyfication and de-fuzzyfication layers correspond to input and output layers in standard NN terminology. Similarly, the AND and OR layers are two hidden layers. There are no more than two hidden layers in the structure, since this would limit the interpretability of the NN, and more importantly it appears redundant since with respect to the Kolmogorov theorem (Atkinson *et al.*, 1997), a NN with one hidden layer can model any function regardless of complexity.

Fuzzyfication Layer

The fuzzyfication layer accepts the numerical data and transforms it to information granules based on fuzzy sets methods. The objective of this process, termed "fuzzyfication," is to find suitable fuzzy sets that best describe the data in each input dimension. Obviously, we have to choose the number of fuzzy sets that best describe our data and also to assign the linguistic tags that best reflect the content. The fewer fuzzy set tags there exist, the more interpretable the results (Nauck, 2003). The choice of linguistic tags is usually straightforward. Valid labels can be any meaningful tags, roughly representing a data granule in this knowledge domain (Nauck, 2003). Another choice related to the fuzzy sets is the approximation function. This is data-dependent and common choices include bell-shaped, triangular, and trapezoidal functions (Bortolan *et al.*, 2002; Dixon, 2004). Dixon (2004) has performed a sensitivity analysis in the hydrology domain, considering fuzzy set shape, number and other factors. Furthermore, an optional choice is fixing an overlapping extent between adjacent fuzzy sets. It is believed that 50 percent overlap produces an error-free encoding (Pedrydz, 2001; Pedrycz *et al.*, 2005). A good practice in creating data justifiable sets is to create a diagram that shows commutative values, preferable normalized, and the fuzzy sets on top of them. A more sophisticated approach is fuzzy equalization (Pedrydz, 2001).

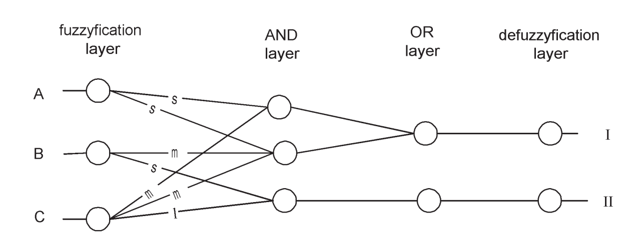

Figure 1. Conceptual skeleton of a simple GNN. A, B, and C are the input dimensions. I and II are the two output classes. Small letters on connections correspond to weights (i.e., small, medium, and large fuzzy sets).

Hidden AND Layer

The nodes of the AND layer (Figure 1) correspond to the number of rules in the rule set. Links connect the fuzzyfication nodes to the AND nodes when the input dimensions corresponding to the particular fuzzyfication nodes are members of the particular rule. Otherwise, the links do not exist, and therefore the network is not fully connected. The weights assigned to GNN links are fuzzy sets rather than the standard small numerical values used in standard NNs.

Hidden OR Layer

The OR layer contains one node per output class. The role of this layer is to capture the rules that lead to the same consequence (i.e., land-use class). The selection is based on the logical operation OR (Pedrydz *et al.*, 2005). An alternative operator, used here, is MAXIMUM (Nauck, 2003). This operator dictates that the rule producing maximum activation function will be selected, and the rest will be neglected. Although those links can have weights, it is often decided to set the weights to one; meaning that no weighting is applied. This is done in order to maintain transparency since weighted rules are hard to interpret (Nauck, 2003). Perhaps it is useful to think of AND and OR neurons and layers in a graphical way as explained by (Lippmann, 1987) where OR nodes join decision hyperboxes formed by sides that correspond to AND nodes.

Defuzzyfication Layer

The output represented by a fuzzy set sometimes needs to be transformed into a scalar value, a task known as the defuzzification process (Song *et al.*, 1994). Mathematically, defuzzification can be treated as a mapping from a high-dimensional space to a lower dimensional space (Song *et al.*, 1994). Each node on the defuzzyfication layer receives an activation in the [0, 1] range. A common strategy is to assign the corresponding category to the node with the maximum value. Selection is in other words done based on the winner-takes-all rationale. The use of other rules that preserve the activation of all output classes is also possible, resulting in nonlinear spectral unmixing systems much like in their standard NN counterparts (Foody *et al.*, 1997; Foody, 2000). Finally, there are several available coding schemas and the right selection can be tricky since it is believed that NNs exhibit chaotic behavior (Kanellopoulos *et al.*, 1997).

Training

Training is based on the back-propagation (BP) algorithm (Rumelhart *et al.*, 1987; Werbos, 1974). As with standard BP, the output error is calculated after each iteration, and a correction is applied to weights according to a predetermined learning rate. One has to keep in mind that with GNNs, it is not initially random numerical weights that change but parameters of the specified fuzzy sets. The difference to standard BP is that learning is not done using gradient decent. A simpler technique is used that directly alters the fuzzy set parameters (Nauck, 2004; Nauck *et al.*, 1999; Wang *et al.*, 1992). It is a more structure oriented approach and a way to see it as a parallelepiped classification using hyperboxes with fuzzy borders. The shape of the hyperbox is progressively altered during training by moving the sides until the performance criterion is met. As with the standard BP, this can be the classification error of the training or testing set, the maximum number of epochs or any other condition.

Training in GNN includes pruning, which is reducing the NN skeleton in terms of nodes and links. Other than promoting transparency, the purpose of skeleton compression is to increase accuracy, forming a network more able to generalize and less prone to overfitting the training data. The removal of skeleton components is done based on a metric of performance. The rule base size can be set to a maximum, meaning that rules are created until all training patterns are covered by rules (Nauck, 2004). On the other hand, an upper limit on the number of rules can be set in order to maintain the interpretability of the network. It is often the case that only a few rules are needed to classify the majority of the samples, but after that, the number of rules to achieve higher accuracy increases dramatically. Starting from the maximum number, pruning progressively removes rules that lead to correct classification less frequently while the overall accuracy of the classification does not drop.

Case Study

Area and Data

The study area selected is located in the island of Lefkas in the western part of the Hellenic Republic and covers approximately 40 km by 40 km. Three sources of information were available for this area: satellite imagery, topographic data, and land-cover data. Satellite imagery comprise one Landsat ETM+ scene acquired on 05 July 2000 from path 185 and row 033. The image consists of seven spectral bands. The topographic data used is a Digital Elevation Model (DEM) prepared under the Shuttle Radar Topography Mission (SRTM, 2000) and made available from the U. S. Geological Survey. More on the accuracy of this dataset in the Hellenic Republic can be found in Miliaresis *et al*. (2005).

Finally, CORINE land-cover was used as reference data. CORINE is a database that currently covers 30 European countries with a common classification scheme and was finalized in the early 1990s as part of the European Commission program to COoRdinate INformation on the Environment (CORINE). A revised version for 2000 is also now available which is used in this study. The classification scheme exhibits three hierarchical levels. Level 1 has just five classes: Artificial surfaces, Agricultural areas, Forest and semi natural areas, Wetlands, and Water bodies. Level 2 has fifteen classes, and Level 3 has forty-four classes. A 150 m positional accuracy and 25 ha minimum mapping unit are stated in the specifications.

The three datasets are georeferenced in a common coordinate system so that they overlap. The geographic projection (WGS84) window coordinates are: lower left 38°32′, 20°30′ and upper right 20°50′, 38°52′. Co-registration error is less than two pixels of the satellite image, which is assumed to be adequate. The slope and aspect are derived from DEM. One can think of the dataset as being an eleven-band image containing the seven bands of Landsat ETM+, together with topographic information and land-cover information. More specifically: B (Blue), G (Green), R (Red), IR (Infrared), MIR5 (Middle-infrared, band 5), T (Thermal-infrared), MIR7 (Middle-infrared, band 7), D (DEM), S (Slope), A (Aspect), C (CORINE). The first ten bands are used as input dimensions in the NN, while the final one (CORINE) contains information about the output class.

As in standard NNs, there is the need to code the output class. The simple binary schema is selected with binary codes corresponding to land-use classes as shown in Table 1. The CORINE land-cover for the study area is shown in Figure 2. Furthermore, all input variables are normalized to the [0,100] range.

Training and Testing Data Set

Fifty samples per class per set were collected, which is the minimum suggested by Congalton (1991). There are no cases with missing values. The CORINE land-cover is used as reference data to determine the output classes. The samples are taken from stratified regions to make sure all classes are covered (Figure 2). The training and testing samples have the characteristics shown on Table 2. The correlation matrix of input parameters is given in Table 3.

Table 1. CORINE Land-cover Level 1 Classes and Codification Used

CODE	Name	CODE NAME
10000	Artificial surfaces	ART
00001	Agricultural areas	AGR
00100	Forest and semi natural areas	FOR
01000	Wetlands	WET
00010	Water bodies	WAT

The CORINE level 1 classification scheme is quite general, containing only five categories. It is selected here to promote comparisons since it is considered as a standard in the European Union. Moreover, it is our intention to make further experiments for level 2 and level 3 in the future. The reader should note, however, that to label the samples, we used CORINE rather that direct ground measurements. This is justified by the fact that highly trained human photo-interpreters, with knowledge of the ground, did the visual interpretation of the satellite images to construct the CORINE land-cover classes. Given the wide adoption of this dataset in all European countries, it seems highly unlikely that CORINE deviates much from what is actually on the ground within its specifications. This argument seems to be even stronger for the specific, relatively simple, five classes used here. Although some of the mis-matches in the accuracy matrices that follow may be attributed to CORINE, this cannot be extensive. The major benefit of using the CORINE data set over any *ad hoc* sampling is the promoted applicability. Using standard data sets such as SRTM and CORINE, the method is readily applicable to a very broad geographical area. We believe that although the use of CORINE does not totally guarantee the accuracy of the numbers presented in this section, it is a reasonable compromise, at least comparable to the use of conventional maps or aerial photographs to obtain the output classes.

Results

Testing Several Configurations

The objective is dual, to find an accurate classifier and at the same time keep the number of rules as low as possible. Approximately one hundred configurations were tested in order to assess the behavior of the system. Most of them are reported here to enable comparisons. Several parameters are kept constant because it is found that they represent near

Table 2. Statistical and Spatial Information

Variable	Mean	Std. Deviation	Min	Max	Spatial Resolution (m)
B	15.99	10.36	4	77	30
G	18.26	12.09	1	81	30
R	24.08	16.52	1	98	30
IR	28.29	15.61	1	69	30
MIR5	37.77	22.01	4	100	30
T	43.39	20.37	7	78	60
MIR7	26.27	16.55	4	86	30
DEM	16.92	21.63	0	93	90
SLOPE	12.96	15.17	0	56	90
ASPECT	38.42	31.60	0	100	90

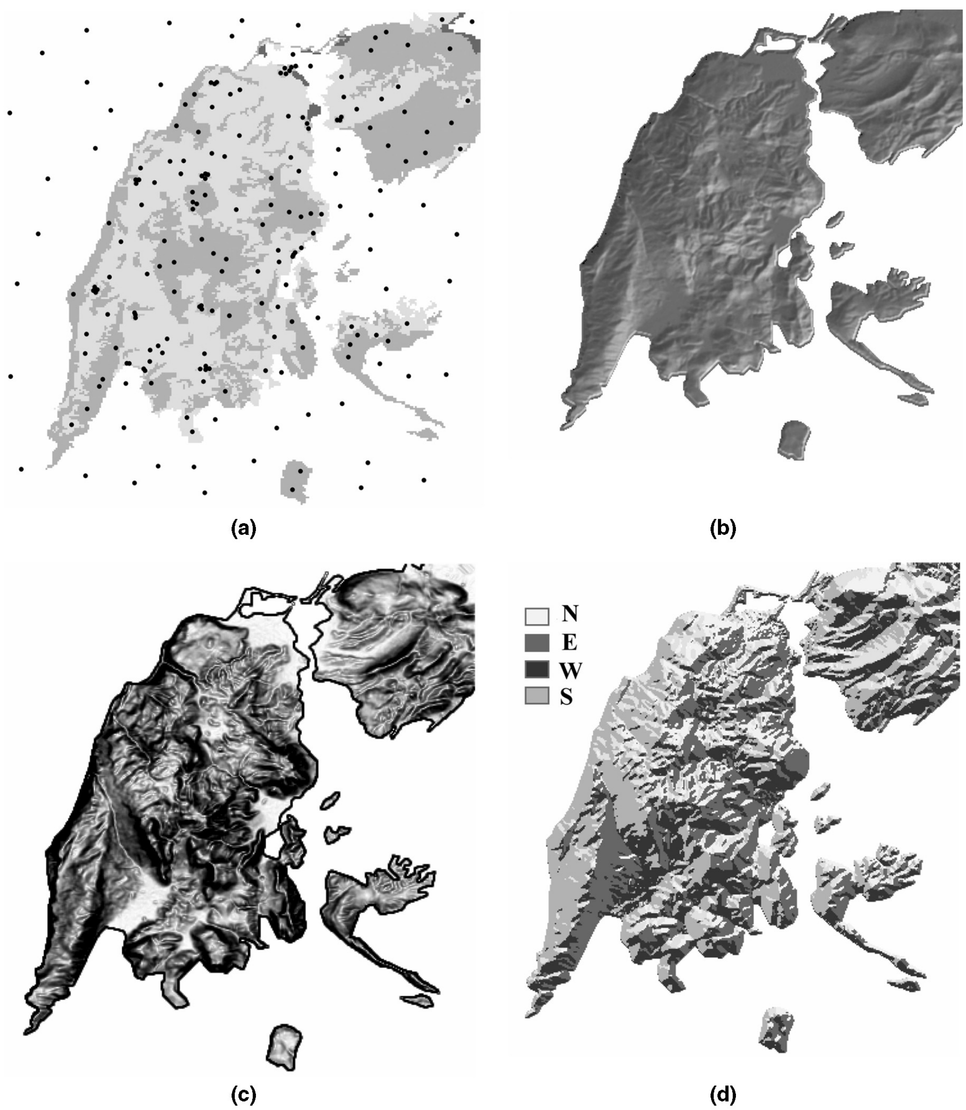

Figure 2. Study area and ancillary data: (a) CORINE level 1 with samples (all classes), (b) Shaded relief derived from the DEM, (c) Slope derived from the DEM, and (d) Aspect derived from the DEM.

Table 3. Correlation Matrix of Input Parameters

	B	G	R	IR	MIR5	T	MIR7	DEM	SLOPE	ASPECT
B	1									
G	0.95	1								
R	0.93	0.99	1							
IR	0.52	0.72	0.74	1						
MIR5	0.68	0.83	0.88	0.92	1					
T	0.32	0.55	0.61	0.85	0.86	1				
MIR7	0.8	0.9	0.94	0.84	0.97	0.75	1			
DEM	0.25	0.37	0.38	0.5	0.55	0.58	0.49	1		
SLOPE	0.22	0.31	0.29	0.38	0.38	0.41	0.33	0.43	1	
ASPECT	0.05	0.21	0.23	0.44	0.37	0.45	0.28	0.27	0.3	1

optimal values since alterations do not yield significant differences. Hence, the maximum number of epochs (training cycles) allowed is set to 2000. The fuzzy sets are forced to overlap but not to a predetermined extent. The relative order of fuzzy sets is kept constant since interpretability would be otherwise spoiled. The winner-takes-all (WTA) rule is used in all runs while in the OR layer the "maximum" operator is used. Pruning aims to produce the best rules per class rather than the best rules overall. This seems to be the only valid option in the land-cover classification context given the

Table 4. Overall Percentage Accuracy (Middle) in Relation to Number of Rules Generated (Bottom) by Changing Configurations (Top). Parameters Kept Constant Include Maximum Number of Iterations, Two Pass Cross validation, and Overlapping Fuzzy Sets. Trapezoidal (trap) and Triangular (triang) Fuzzy Sets are Shown. Names of Inputs are R(Red), IR(Infrared), T(Thermal Infrared), D(DEM), S(Slope), A(Aspect)

Fuzzy Set Number	10	5	3	5	5	5	5	5	5	5
Learning Rate	0.05	0.05	0.05	0.1	0.075	0.05	0.025	0.01	0.001	0.0001
Fuzzy Set Form	trap	triang	trap	trap	trap	trap	trap	trap	trap	trap
ALL SPECTRAL	80	69	63	72	74	72	71	74	75	74
ALL	96(a)	82	74	88	84	81	80	80	78	78
R –IR-T-D	85	78	70	77	78	78	77	77	77	77
R –IR-T-D S	91	78	71	79	79	80 (b)	80	76	76	77
R –IR-T-D-S-A	95	79	74	80	82	83	82	76	77	78
R –IR-T-S	83	70	68	76	77	76	76	76	76	76
ALL SPECTRAL	63	21	8	33	16	32	10	27	26	25
ALL	162 (a)	70	34	90	70	53	60	54	57	57
R –IR-T-D	66	19	8	21	16	14	12	12	12	12
R –IR-T-D S	101	33	13	32	16	24 (b)	15	14	16	16
R –IR-T-D-S-A	151	57	23	80	64	55	44	40	41	43
R -IR-T-S	72	20	10	26	20	17	17	17	17	17

requirement that all classes are present in the output. There is no minimum acceptable error percentage or minimum number of epochs predefined. The network is trained for ten more iterations after the performance criterion is met. For the validation of the system, the training set is divided into two equal sized sets. In a first pass, the first set is used for training and the second for testing, and then the sets are swapped in a second pass. Accuracy is reported as the average of the two passes. We refer to this as two pass cross validation.

Table 4 is divided in three sections. The top section contains ten different configurations. The middle section contains the overall accuracy achieved in the classifications by using six different input vector combinations. The rows of the middle section correspond to (a) all seven spectral bands with no ancillary information (b) all seven spectral and ancillary information (DEM, slope, and aspect), resulting in ten inputs to the fuzzification layer, (c) Red, Infrared, Thermal, and DEM only, (d) Red, Infrared, Thermal, DEM, and Slope, (e) Red, Infrared, Thermal, DEM, Slope, and Aspect, and (f) Read, Infrared, Thermal, and Slope. The same six input combinations are repeated in the rows of the bottom section of Table 4 that shows the number of rules generated for each configuration. Taking as an example the underlined values marked as (a) in the table it should be read as: A GNN with 10 Fuzzy sets, 0.05 Learning rate and trapezoidal approximation function (top) yields 96 percent accuracy (middle) and results to 162 rules (bottom). Another example is the underlined one marked (b) where five fuzzy sets with 0.05 learning rate and trapezoidal approximation function yield 80 percent overall accuracy, but this time with only 24 rules. Those two configurations are further exploited in the following discussion. The first one is selected as representative of the high accuracy level that a complex network can achieve, but clearly results in a difficult to interpret skeleton 10:162:5:5 (input:AND:OR:output, number refers to the number of nodes). By the number of required rules, we can tell that water is by far the easiest class whereas wetlands is a bit less difficult than the other three. Interestingly, the 162 rules are distributed per category as follows:

$$(42, 32, 44, 3, 41) \rightarrow (\text{ART, WET, FOR, WAT, AGR}) \\ \text{number of rules per class.}$$

The second case, marked (b) in Table 4, is selected as the most accurate network between those that resulted in the smallest set of rules. The input dimensions are the three least correlated spectral bands plus DEM and Slope.

Only one instance of triangular fuzzy sets is finally shown in Table 4. This is because triangular fuzzy sets are found to perform slightly worse than trapezoidal sets when all other parameters are kept constant. Additionally, the computational cost of using trapezoidal over triangular sets is negligible. This statement holds for the particular data set used and further work is required to make a more general evaluation of the impact of the shape of the fuzzy sets. A preliminary sensitivity analysis can be found in Dixon (2005).

Accuracy Assessment

Accuracy assessment is probably the most important aspect of any classification process. It is not possible to improve a classification method unless we establish a procedure to measure the performance. The recent paper by Wilkinson (2005) is important because it is found that there has been no demonstrable improvement in the classification performance over the last 15 years. However, it is found there that a number of recent papers do not report even basic classification performance metrics. As a general accuracy standard, it is assumed here that the overall classification error should be over 85 percent with no class less than 70 percent (Thomlinson *et al.*, 1999).

The accuracy matrix has been proposed as a good basic accuracy statement (Congalton and Green, 1998). One of the important aspects of this matrix is that it shows the distribution of error per class as well as commission and omission errors. This very useful information is masked out when only overall accuracy is reported. The internal distribution of error is not only useful to document accuracy. It is in fact a very effective analytical tool since by examining commission and omission errors, we are able to understand some properties of the training samples chosen. For example it is found at the initial stages of the present study that the agricultural class was confused with the urban class, which is not common given their quite different spectral properties. This led to re-examining the training samples used for the agricultural class and discovering that a few samples were actually located on or near artificial structures such as buildings and roads. Moving those samples a few pixels away led to less confusion in the accuracy matrix.

In addition, the kappa coefficient is used to promote comparison with different classifications. Comparisons are possible because this coefficient takes into account the fact that some samples are allocated to the correct class purely by chance (Congalton and Green, 1998). The following formula is used to calculate kappa coefficient (Congalton and Green, 1998; Foody, 2000):

$$\hat{K} = \frac{n\sum_{k=1}^{q} n_{kk} - \sum_{k=1}^{q} n_{k+}n_{+k}}{n^2 - \sum_{k=1}^{q} n_{k+}n_{+k}}$$

where q is the number of classes, and n is the total number of training samples.

Most Accurate Solution

The results of the run with the highest accuracy are shown in Table 5. There are two cases that the neural network did not classify at all. The first instance, belonging to the wetlands category, is manually assigned to water, and the second instant, belonging to agricultural, is assigned to forest. In both cases all elements in the output vector are zero. The assignments are necessary to construct the accuracy matrix with the provision that the results are not significantly altered.

The "not-classified" case is a result of there being more than one output classes receiving equal activation to the given degree of floating point resolution. The WTA rule cannot handle this case because it cannot decide how to classify the two or more output classes which receive equal output activation. An easy solution to this problem is to select the output class randomly among the selected candidates. Another case of unclassified samples is when zero activation is output for all classes. This can result from forcing the system to have too few rules. The solution is to include more rules and accept some inevitable transparency loss.

Most Transparent Solution with High Accuracy

The run that generates relatively fewer rules and at the same time maintains acceptable level of accuracy is presented in Table 6. The rule set is presented in Table 7. Apparently 24 rules are selected out of the initially 90 rules after pruning. In both cases, as shown in Tables 5 and 6, the overall accuracy percentages are slightly different compared to that reported on Table 4 because the former corresponds to the best run whereas the latter to the average of two runs.

There are four not classified instances with the following target and output vector pairs. A forest case, manually assigned to wetlands with vector:

$$\begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix} \rightarrow \begin{bmatrix}0.000\\0.452\\0.187\\0.000\\0.452\end{bmatrix} \qquad \text{Wetlands vector is} \begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix}.$$

Three agricultural cases manually assigned to agricultural with vectors:

$$\begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix} \rightarrow \begin{bmatrix}0.000\\0.000\\0.423\\0.000\\0.423\end{bmatrix}, \quad \begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix} \rightarrow \begin{bmatrix}0.223\\0.000\\0.000\\0.000\\0.223\end{bmatrix}, \quad \begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix} \rightarrow \begin{bmatrix}0.229\\0.229\\0.158\\0.000\\0.229\end{bmatrix}.$$

Table 5. Accuracy Matrix for the Most Accurate Solution, Case (a) in Table 4 Resulted Nevertheless in a Large Set of Rules

	ART	WET	FOR	WAT	AGR	Σ
ART	49	1	0	0	0	50
WET	1	47	0	0	0	48
FOR	0	0	50	0	1	51
WAT	0	2	0	50	0	52
AGR	0	0	0	0	49	49
Σ	50	50	50	50	50	250

	Producer's Accuracy	User's Accuracy	
ART =	98%	98%	
WET =	94%	98%	overall accuracy = 98%
FOR =	100%	98%	
WAT =	100%	96%	KHAT = 98%
AGR =	98%	100%	

Table 6. Accuracy Matrix for the Most Accurate Solution that Resulted in a Small Set of Rules. Case (b) in Table 4

	ART	WET	FOR	WAT	AGR	Σ
ART	39	5	3	0	4	51
WET	4	43	1	0	1	49
FOR	3	0	39	0	6	48
WAT	0	1	0	50	0	51
AGR	4	1	7	0	39	51
Σ	50	50	50	50	50	250

	Producer's Accuracy	User's Accuracy	
ART =	78%	76%	
WET =	86%	88%	overall accuracy = 84%
FOR =	78%	81%	
WAT =	100%	98%	KHAT = 80%
AGR =	78%	76%	

A Standard NN Classification for Comparison

To make it possible to appreciate more the comparison of the GNNs over conventional NN approaches, we also present the results of a classification using a standard NN. The actual NN used is a feedforward multilayer perceptron with topology of 5:20:5. Training is performed using backpropagation with momentum (0.9) and learning rate equal to 0.03. The Mean Squared Error (MSE) is used as the performance function and the transfer function in both the hidden and the output layer is log-sigmoid. The parameters are set experimentally and the termination criterion is set to MSE $\leq$ 0.04. It is reached after approximately 50,000 epochs, and the results are presented in Table 8.

It is seen that the standard neural network results in a better overall classification accuracy when compared to the transparent GNN but is less accurate to the best GNN. This might suggest that GNNs with unlimited number of rules and for certain data sets can classify more accurately than standard NNs possibly due to information granulation. It also suggests that transparency comes at a cost. However, this cost, that is the difference in overall classification accuracy, seems to be relatively small and for some applications may be insignificant compared to the gain of structure interpretability.

Interpreting the Network

The resulting GNN skeleton corresponding to a 5:24:5:5 topology is presented in Figure 3. Figure 4 presents detailed information related to the agricultural class. Four rules are

Table 7. Rule Set: 24 Rules are Selected Out of the 90 Initial Rules After Pruning

Red	Infrared	Thermal	DEM	Slope	Output Class
small	*medium*	*medium*	*very small*	*very small*	*AGR*
small	*medium*	*medium*	*medium*	*small*	*AGR*
small	*medium*	*medium*	*very small*	*small*	*AGR*
small	*medium*	*large*	*medium*	*small*	*AGR*
medium	*medium*	*large*	*small*	*very small*	*ART*
medium	*medium*	*medium*	*very small*	*very small*	*ART*
medium	*medium*	*large*	*small*	*small*	*ART*
medium	*medium*	*small*	*very small*	*medium*	*ART*
medium	*medium*	*medium*	*very small*	*small*	*ART*
medium	*medium*	*large*	*medium*	*very small*	*ART*
	large	*medium*	*large*	*very small*	*ART*
very small	*small*	*medium*	*small*	*small*	*FOR*
very small	*small*	*medium*	*small*	*medium*	*FOR*
small	*medium*	*large*	*small*	*medium*	*FOR*
medium	*medium*	*large*	*very large*	*medium*	*FOR*
very small	*medium*	*large*	*medium*	*medium*	*FOR*
small	*medium*	*medium*	*very large*	*small*	*FOR*
medium	*medium*	*large*	*very large*	*small*	*FOR*
very small	*very small*	*very small*	*very small*	*very small*	*WAT*
small	*small*	*small*	*very small*	*small*	*WET*
small	*small*	*small*	*very small*	*very small*	*WET*
	medium	*large*	*very small*	*medium*	*WET*
medium	*small*	*medium*	*very small*	*very small*	*WET*
small	*small*	*small*	*very small*	*medium*	*WET*

enough to classify the agricultural class at an accuracy level of approximately 77 percent. This class is governed by low visible red reflectance and medium infrared reflectance. Thermal reflectance is medium to large. It appears that the agricultural land is located in very low to medium altitude, which confirms our general knowledge that high altitudes do not favor the presence of cultivated land. The same finding applies to slope as well. The fuzzy sets for this configuration are five initially symmetrical and evenly distributed over the histogram of input values. After training the fuzzy set parameters are altered as shown in Figure 5. Note that some input dimensions could possibly be represented by less fuzzy sets. In future experiments, it would be worth trying to have different numbers of fuzzy sets per variable to see whether the same accuracy level can be achieved with a smaller number of fuzzy sets, resulting to a more compact rule set.

Discussion

There are several results that could be used as future guidelines in similar studies. The overall capacity of GNNs, having a large number of fuzzy sets (more than seven) and letting the number of rules grow without restriction could result in very high accuracies of over 90 percent. This is a similar finding to standard neural networks where complex structures can classify quite accurately. The drawback is that increased complexity makes it impossible to logically explain or visualize the acquired knowledge.

Concerning the number of fuzzy sets, it seems to be impossible to approach the 85 percent overall accuracy standard with any configuration based on three fuzzy sets. The maximum accuracy achieved with three fuzzy sets is around 70 percent. Using five fuzzy sets is sufficient to reach the 85 percent level. Using seven fuzzy sets usually

Table 8. Accuracy Matrix for a Standard NN Classification

	ART	WET	FOR	WAT	AGR	Σ
ART	40	1	2	0	2	45
WET	2	45	0	0	0	47
FOR	1	1	38	0	9	49
WAT	0	0	0	50	0	50
AGR	7	3	10	0	39	59
Σ	50	50	50	50	50	250

	Producer's Accuracy	User's Accuracy	
ART =	80%	89%	
WET =	90%	96%	overall accuracy = 85%
FOR =	76%	78%	
WAT =	100%	100%	KHAT = 81%
AGR =	78%	66%	

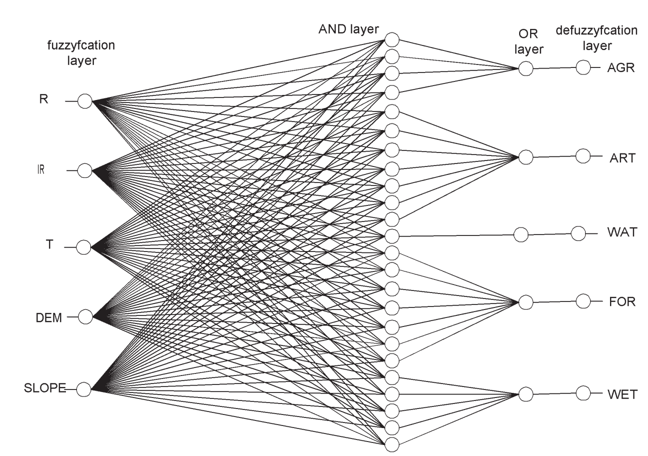

Figure 3. Resulting GNN skeleton after pruning.

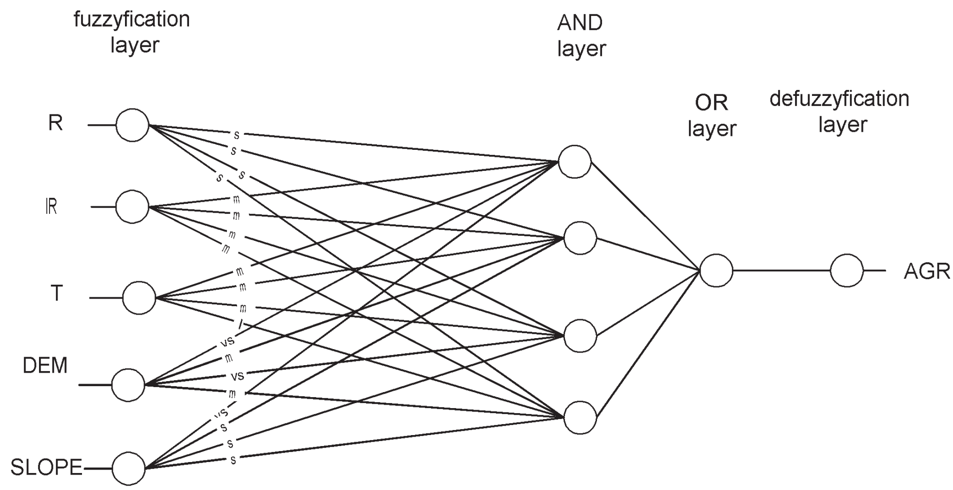

Figure 4. The Agriculture classification path extracted from the neural network skeleton. It consists of four rules. Letters on links denote weights: vs, s, m, l, vl that correspond to very small, small, medium, large, and very large (not present here), respectively.

yields a marginal increase compared to the five fuzzy sets for this dataset.

With regard to the form of the approximation function that represents the fuzzy sets, it seems that trapezoidal sets are marginally better than triangular sets. Bell shaped sets fail completely which has been confirmed with other studies (Dixon, 2004). The reason remains unknown since in principle a bell shaped approximation should at least be suitable to model phenomena that exhibit Gaussian distribution.

We believe that the number of epochs required for the training, it is not directly related to the performance unless the maximum number of sets is too low causing premature convergence. Often a qualified solution is found before the maximum number of epochs (2,000) is reached. Even when training exhausts the pre-specified number of iterations, accuracy gain is minor because error initially drops rapidly and then descends very slowly.

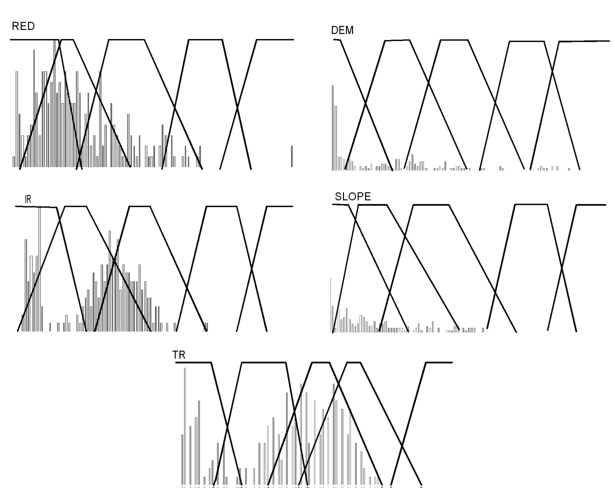

Figure 5. Fuzzy sets after training overlaid on the normalized cumulative distribution of digital numbers. Fuzzy sets are very low, low, medium, large, and very large from left to right.

Pruning, that is, reducing the number of rules to an interpretable size always comes at a cost which is a loss in accuracy. Limiting maximum rules to ten yields an accuracy level of 70 percent. The number of rules required to achieve a higher accuracy level is then logarithmically increased. Instead of limiting the maximum number of rules, it is possible to keep the skeleton manageable by having just three fuzzy sets, which also produces approximately 70 percent overall accuracy.

Learning rate proves to be an important factor. Keeping all other parameters fixed while changing the learning rate produces a variation in accuracy as much as 10 percent. The impact on rule set size is even more profound since some learning rate values cause the reduction of rules in the set to as much as 50 percent. The addition of a momentum as in standard BP algorithms might be recommended.

An efficient way to reduce complexity is to decrease input dimensionality, but the dimensions that are kept should be carefully selected. Preserving only the spectral bands, results in low overall accuracy (70 percent) with unbalanced internal distribution of error. A more efficient structure is formed by keeping the less correlated spectral bands with the addition of DEM and slope data sets. The inclusion of aspect increases accuracy, but simultaneously causes an undesirable increase in the complexity of the NN. It is our intention to use a more structured feature selection method (Ruck *et al.*, 1999) in the future.

Finally, further research is required on the prefixed overlapping extent between adjacent fuzzy sets. The fact that there is no fixed overlapping could be one more reason for the presence of unclassified instances. Making sure that all samples are classified is desirable, so that an accuracy matrix can be constructed. The 50 percent overlapping could be adopted provided that it eliminates unclassified instances, and at the same time it does not negatively affects classification in terms of accuracy or transparency.

Conclusions

We have shown that GNN techniques can be used with satellite data in the land-use classification context to yield very accurate or meaningful results. High accuracy comes at the cost of transparency, because the number of rules dramatically increases. The most interpretable results, with just a few rules, are not characterized by as high classification accuracies. Clearly, there is a trade-off between accuracy and interpretability. In addition, GNNs offer a suitable paradigm for the inclusion of ancillary information in the classification process. In this study, topographic ancillary information was employed but other types may also be used. Indeed, we plan to extend the method to CORINE levels 2 (medium scale) and 3 (urban). In urban scale and using VHR satellite imagery the incorporation of other types of ancillary information is promising to separate difficult classes such as dump and construction sites.

Finally, GNNs offer two very interesting characteristics that can be further investigated in the future. One, already mentioned previously, is spectral unmixing. The other is the use of symbolic input variables rather than numeric ones (Nauck, 2004). For example it would be more appropriate to code aspects in the present study as W, E, S, and N instead of having the actual bearing values. Another example would be the use of a soil map as ancillary information.

Acknowledgments

The three anonymous referees are gratefully acknowledged for making comments which substantially improved the manuscript. We would also like to thank Paul Clausen and Linda Stephen for their help in preparing the final manuscript. The intellectual property rights of the satellite image used belong to IMAGE2000 of JRC based on Landsat 7 ETM+ © ESA distributed by Eurimage; ortho-correction EU15 © Metria. CORINE is © EEA, Copenhagen, 2000 and may be downloaded from **http://dataservice.eea.eu.int/dataservice/metadetails.asp?id=759**.

References

Atkinson, P.M., and A.R.L Tatnall, 1997. Introduction neural networks in remote sensing, *International Journal of Remote Sensing*, 18(4):699–709.

Benediktsson, J.A., P.H. Swain, and O.K. Ersoy, 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data, *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):540–552.

Benediktsson, J.A., J.R. Sveinsson, and P.H. Swain, 1997. Hybrid consensus theoretic classification, *IEEE Transactions on Geoscience and Remote Sensing*, 35(4):833–843.

Benitez, J., J. Castro, and I. Requena, 1997. Are artificial neural networks black boxes?, *IEEE Transactions on Neural Networks*, 8(5):1156–1164.

Bortolan, G., 1998. An architecture of fuzzy neural networks for linguistic processing, *Fuzzy Sets and Systems*, 100:197–215.

Bortolan, G., and W. Pedrycz, 2002. Linguistic neurocomputing: The design of neural networks in the framework of fuzzy sets, *Fuzzy Sets and Systems*, 128:389–412.

Congalton, R.G., 1991. A review of assessing the accuracy of classification of remotely sensed data, *Remote Sensing of Environment*, 37:35–46.

Congalton, R.G., and K. Green, 1998. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Lewis Publishers, 137 p.

Dixon, B., 2005. Applicability of neuro-fuzzy techniques in predicting ground-water vulnerability: A GIS-based sensitivity analysis, *Journal of Hydrology*, 309:17–38.

Foody, G.M., 2002. Status of land cover classification accuracy assessment, *Remote Sensing of Environment*, 80:185–201.

Foody, G.M., 2000. Estimation of sub-pixel land cover composition in the presence of untrained classes, *Computers & Geosciences*, 26:469–478.

Foody, G.M., R.M. Lucas, P.J. Curran, and M. Honzak, 1997. Nonlinear mixture modelling without end-members with an artificial neural network, *International Journal of Remote Sensing*, 18:937–953.

Kanellopoulos, I., and G.G. Wilkinson, 1997. Strategies and best practice for neural network image classification, *International Journal of Remote Sensing*, 18(4):711–725.

Lee, S.C., and E.T. Lee, 1975. Fuzzy neural networks, *Mathematical Biosciences*, 23:151–177.

Lippmann, R.P., 1987. An introduction to computing with neural nets, *IEEE ASSP Magazine*, 2:4–22.

Miliaresis, G.Ch., and Ch.V.E. Paraschou, 2005. Vertical accuracy of the SRTM DTED level 1 of Crete, *International Journal of Applied Earth Observation and Geoinformation*, 7:49–59.

Nauck, D., 2004. Neuro-fuzzy learning with symbolic and numeric data, *Soft Computing – A Fusion of Foundations, Methodologies and Applications*, 8(6):383–396.

Nauck, D., 2003. Fuzzy data analysis with NEFCLASS, *International Journal of Approximate Reasoning*, 32:103–130.

Nauck, D., and R. Kruse, 1999. Neuro-fuzzy systems for function approximation, *Fuzzy Sets and Systems*, 101:261–271.

Pedrydz, W., 2001. Fuzzy equalization in the construction of fuzzy sets, *Fuzzy Sets and Systems* 119:329–335.

Pedrydz, W., and G. Vukovich, 2001. Granular neural networks, *Neurocomputing*, 36:205–224.

Pedrycz, W., and M. Reformat, 2005. Genetically optimized logic models, *Fuzzy Sets and Systems*, 150:351–371.

Pedrycz, W., and A. Vasilakos, 1999. Linguistic models and the linguistic modeling, *IEEE Transactions on Systems, Man, Cybernetics, B*, 29(6):745–757.

Ruck, D.W., S.K. Rogers, and M. Kabrisky, 1999. Feature selection using a multilayers perceptron, *Journal of Neural Network Computing*, 2(2):40–48

Rumelhart, D.E., G.E. Hinton, and R.J. Williams, 1987. Learning internal representations by error propagation, *Parallel Distributed Processing-Explorations in the Microstructure of Cognition* (D.E. Rumelhart, J.L. McClelland, PDP Research Group, editors). The MIT Press, Boston, Massachusetts, 1:318–362.

Shuttle Radar Topography Mission (SRTM), 2000. URL: http://www2.jpl.nasa.gov/srtm/ (last date accessed: 17 October 2007).

Song, Q., and G. Bortolan, 1994. Some properties of defuzzification neural networks, *Fuzzy Sets and Systems*, 61:83–89

Thomlinson, J.R., P.V. Bolstad, and W.B. Cohen, 1999. Coordinating methodologies for scaling landcover classification from site-specific to global: Steps toward validating global map products, *Remote Sensing of Environment*, 70:16–28.

Wang, L.X., and J.M. Mendel, 1992. Generating fuzzy rules by learning from examples, *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6):1414–1427

Werbos, P.J., 1974. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, Ph.D. thesis, Harvard University, Boston, Massachuttes.

Wilkinson, G.G., 2005. Results and implications of a study of fifteen years of satellite image classification experiments, *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):433–440.

Zadeh, L.A., 1965. Fuzzy sets, *Information and Control*, 6(8):338–353.

Zadeh, L.A., 1999. From computing with numbers to computing with words – From manipulation of measurements to manipulation of perceptions, *IEEE Transactions on Circuits and Systems – I*, 45(1):105–119.

Zadeh, L.A., 1997. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems*, 90:111–127

(Received 09 February 2006; accepted 12 April 2006; revised 07 June 2006)