Neural networks analysis of satellite images for land cover discrimination

F. CAPARRINI, E. CAPORALI, F. CASTELLI

Department of Civil Engineering, University of Florence, Via S.Marta 3, I-50139 Florence (Italy).
e-mail: fcapa@dicea.unifi.it Tel (+39) 055 4796317 Fax (+39) 055 495333

ABSTRACT

An adequate knowledge of land cover characteristics is very important for hydrological and meteorological modelling. Unfortunately, available maps are often out of date, since land use can change considerably even in short periods. Therefore, in the recent years there has been an increasing interest in the use of satellite images for the discrimination and mapping of land cover, based on the different spectral response patterns of the various types of earth surface. In the present work two *Landsat Thematic Mapper* images have been used to obtain land cover maps of the Arno basin, one of the major watersheds in Central Italy. The images have been processed using both "classical" techniques like a Bayesian-Maximum Likelihood classifier and Cluster Analysis, widely used in remote sensing applications, and with a more innovative approach with Neural Networks. The advantage of using Neural Networks for image classification is that such techniques appear to be more "flexible" than Bayesian classifiers, since they do not require any a priori assumption about the class statistical distribution in the data set. Different Network architectures have been trained and applied and the resulting thematic maps have been compared with ground truth data and available cartography. Also, different levels of discrimination have been tested, including the recognition of vegetation types.

1 INTRODUCTION

In the recent years there has been an increasing interest in the use of Artificial Neural Networks to classify remotely-sensed data (Bishop *et al.*, 1992, 1998; Fardanesh and Ersoy, 1998; Yoshida and Omatu, 1994) and for other hydrologic applications (Castellani *et al.*, 1996).

In several cases, such methods have been demonstrated to give better results than the "traditional" classifiers like Bayesian-Maximum Likelihood rule or unsupervised techniques (Cluster Analysis).

This paper briefly describes the processing techniques that can be used to obtain land cover maps from satellite images, with special emphasis on neural network procedures.

The main advantage in using Neural Network classifiers is that they do not require any *a priori* assumption in the classes statistical distribution, since they are non-parametric classifiers. Furthermore, the ability of Neural Networks to "learn" and adapt to different situations makes them more flexible and potentially capable of recognizing also inputs with higher degrees of noise.

In the present study two Landsat Thematic Mapper images, taken in 1991 in the Arno basin, have been classified with three different methods: Maximum Likelihood, Cluster Analysis (unsupervised classification) and Neural Networks.

The area covered by the images (92.5 x 80 *km*) is the South-Western part of Tuscany. Therefore, only the lower part of the Arno basin is included in the images. The resulting study area, obtained superimposing the "mask" of Arno basins on the two TM quarter-scenes, turns out to be made of 4290333 pixels, corresponding to an area of 3860 square kilometers. The images were rectified and georeferenced in UTM projection by the identification of 204 Ground Control points.

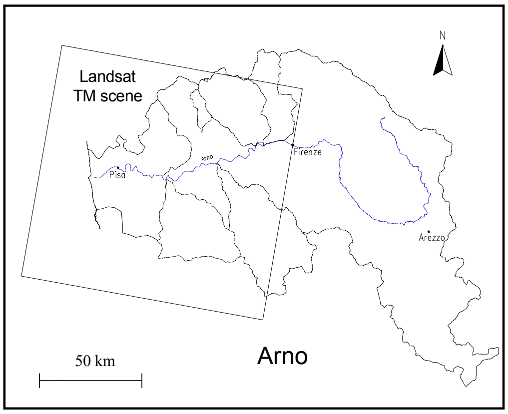

Figure 1: The study area.

The data set used as ground truth is taken from a field campaign carried out for Regione Toscana in 1993 (IFT Regione Toscana, 1996). It consists in direct surveying of one point every 400 *m*. The original data set has 69 classes.

It should be noted that some classes are separated in the legend but actually represent the same kind of land cover, in fact they differ mainly for "socio-economic” aspects and much less for surface characteristics (for example, the classes “residential urban areas”, “commercial urban areas” etc.). On the other hand, the class “cultivated land” actually includes many different surface types (depending on the type and method of farming). The legend in the land cover maps obtained from satellite data will then be obviously different in some aspects from the reference one.

2 MULTISPECTRAL CLASSIFICATION

The *classification* procedures aim to automatically categorize all pixels in a satellite image into land cover classes.

The physical basis on which all procedures rely is that different kind of Earth surfaces have different spectral reflectance and emittance properties. For example, figure 2 shows the typical reflectance curves of water, bare soils and vegetation. Similar surfaces will have similar curves, with differences related to intrinsic characteristics like moisture content, organic content (for soils), foliage distribution and structure, biomass (for vegetation), presence of sediments and pollutants (for water).

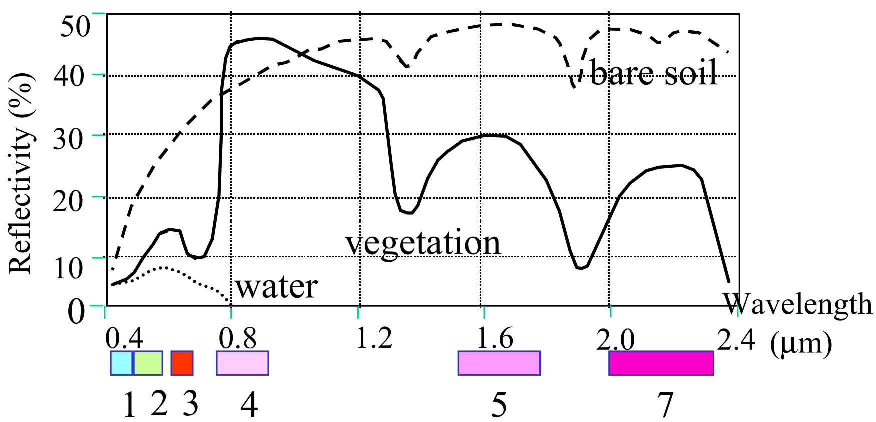

Figure 2: Spectral response patterns of water, bare soil and vegetation and position of the Landsat Thematic Mapper bands.

In Landsat TM data, values of reflectance are converted in Digital Numbers (DNs) between 0 and 255 (8 bits). Therefore, different kinds of surfaces will present different combinations of DNs. By analyzing these combinations of DNs, it is possible to individuate the pixels that have similar characteristics and assign them to a given class.

Generally it is advisable to use multi-temporal images, so that the discrimination is improved due also to the seasonal variation of vegetated surfaces (for example evergreen and deciduous vegetation can be separated).

Then, the classification algorithms take as input the values of reflectance DNs in the 6 bands of winter image and 6 bands of the summer one (note that the thermal band - TM6 - is usually not used in classification procedures), and outputs a thematic map where each pixel is assigned to a specific land cover class.

“Traditional” methods used in satellite image processing can be divided in *unsupervised* and *supervised* techniques.

2.1 Unsupervised Classification

Unsupervised techniques are based on Cluster Analysis. They use multivariate statistical procedures in order to find “natural groupings” of pixels in an image, i.e. spectrally separable classes. The basic premise is that values within a given land cover type should be close together in the measurement space, whereas data in different classes should be comparatively well separated.

The procedure is automatic, based only on the spectral values, with no *a priori* assumption; thus the identity of each class is not initially known, but it has to be assigned *a posteriori* after the classification.

The results obtained depend on the input variables and on the clustering procedure applied.

In the present case, a Cluster Analysis has been performed through ISODATA algorithm (Iterative Self Organising Data Analysis Technique), an iterative process based on the spectral distance (Euclidean) between pixels (Tou and Gonzales, 1974).

The algorithm requires to specify the number of classes, the maximum number of iterations and the threshold value. These parameters have been set respectively to 69 (classes, like the reference ground truth data), 12 (max iterations) and 0.99 (threshold value, meaning that if after an iteration the 99% of pixels is classified in the same group of the previous one, the process automatically stops). In our case, the process stops after 12 iterations, having reached a convergence of 0.97).

In order to identify the 69 classes obtained applying the ISODATA procedure, we have simply calculated the cross-statistic between the clustered image and the reference data, i.e. we have calculated the distribution of ground truth data into every Cluster. Then, we have assigned to each cluster class the land cover type who has the majority of values.

2.2 Supervised Classification

In *Supervised Classification*, the image analyst "trains" the pixel categorization process by specifying representative sample sites of known land cover type (*training areas*). Such samples are used to obtain numerical interpretation keys that describe the spectral attributes of each class.

Each pixel in the image is then assigned to a spectral class according to an appropriate decision rule defined by the analyst.

The steps to follow in Supervised Classification are:

- Definition of the Legend, i.e. which land cover classes are supposed to be in the image to classify
- Individuation of one or more training areas for each class in the legend
- Choice of a decision rule and evaluation of the training set
- Classification

As regard the decision rule, the most reliable one, and the most used, is the *Maximum Likelihood* Algorithm. This decision rule is based on the probability that a pixel belongs to a particular class. For a detailed description, see Lillesand and Kiefer (1987), Rees (1990).

The basic equation assumes that these probabilities are equal for all classes, and that the input bands have normal distributions. For this reason, the training samples that turn out not to have a normal distribution can't be used for classification.

Since it was impossible to obtain for all the 69 classes samples with a normal distribution, in this study we have classified the Landsat image using a simplified legend that has only 11 classes.

3 ARTIFICIAL NEURAL NETWORKS

A Neural Network model is composed by simple and highly interconnected processing units (neurons), which send signal to one another with appropriate *weights* (w_{ij} denotes the weight of the connection from unit i to unit j). A network unit has a rule for summing the signals coming in and a rule for calculating an output signal that is sent to the other units. The rule for calculating the output is known as the *activation* or *transfer function*.

A neural network can be trained to associate an output to a given input. This is obtained by giving a set of training inputs and the corresponding targets. During training, the weights change according to a specified rule until they reach an 'optimal' value.

To perform the classification with Neural Networks, the DNs (0-255) values of 12 TM bands (all bands except the thermal one for the two images) for each pixel are presented as input. The target classes are represented in vector form i.e. they are represented by vectors (1 x number_of_classes) in which the activation of unit *i* is set to 1 when class *i* is being represented and all other units are set to 0. (For example, target for class 2 is [0 1 0 0 0 0 0…] and so on).

The network scheme adopted is a 2-layers feed-forward (*back-propagation*) network, both layers with a **log-sigmoid** transfer function (the log-sig function is particularly suitable because its output ranges from 0 to 1).

The back-propagation algorithm is the most used in multilayer Neural Networks. It defines two sweeps of the network: a forward one from the input layer to the output layer, and then a backward one when the errors are back-propagated to determine how the weights should be changed during training. Learning is achieved using a generalization of the delta rule. A full description of the back-propagation algorithm can be found in any basic Neural Network book like Callan (1999).

The steps to follow are:

- *Initialization*: The network is created specifying the number of layers, the number of neurons in each layer and the transfer function -in this case log-sigmoid:

$$f(net_i) = \frac{1}{1 + \exp(-net)_i}$$

- *Training*: Given the input test-vector (i.e. the values of reflectivity in the 12 TM bands) and the target vector (the ground truth data, presented in vector form as described above), the network is trained until either the error (measured summed squared error) has reached an acceptable minimum (see figure 3), or a maximum number of iterations is reached, or the gradient has approached a minimum level. All the training parameters can be set by the analyst. Several functions can be used to train feed-forward networks: In our case, the fastest one proved to be a gradient-descent with momentum (*traingdm*).

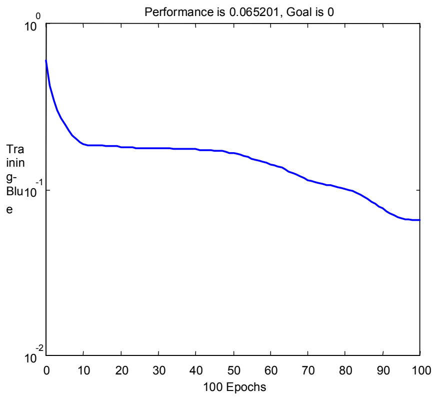

Figure 3: Error plot during 100 epochs of training.

- *Simulation:* Once the network has been trained on the pixels that represent the set of ground-truth data, it is applied to the full image. The output is then recorded from vector to index form. In order to obtain a vector which is composed by 1 (for target class) and 0, it is necessary to apply a competitive function, that is a function that outputs 1 where the net input vector has its maximum value and 0 elsewhere. Finally, the vector obtained is reshaped to the image size and a map of the classified data is produced.

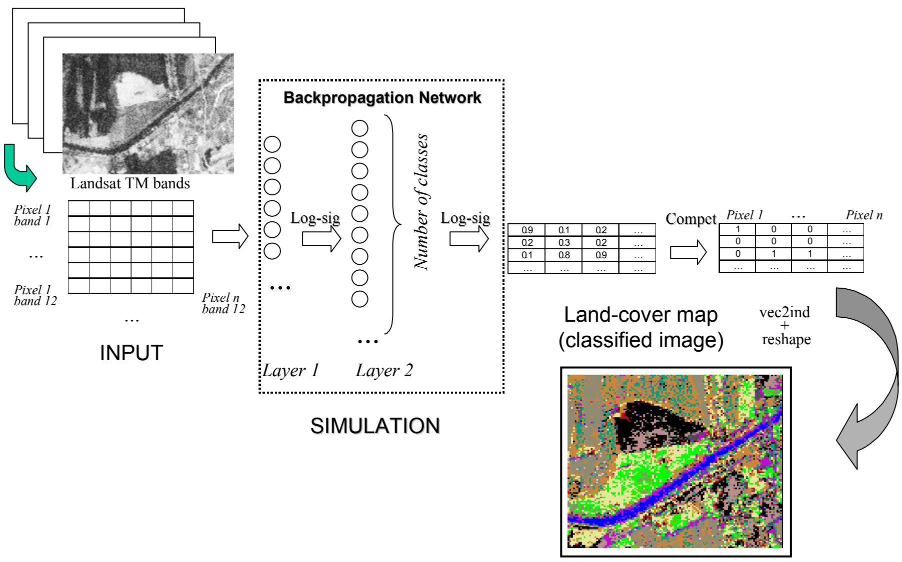

Figure 4: Classification of Landsat images with a feed-forward Neural Network.

3.1 Network architecture

The number of neurons in the first layer (also called hidden layer) can be set arbitrarily. With a higher the number of hidden neurons, the network will be better capable of recognizing the input vectors; on the other hand, the training time will be longer. As regards the second layer (the *output layer*) the number of neurons is determined by the number of classes to recognize; the layer will have as many neurons as the classes in the legend.

We have set the number of hidden neurons to 100; and for the second layer we have tried different numbers of output classes:

- A legend with 69 classes, i.e. all the classes in the original ground truth dataset. In this case, due to the big amount of ground truth data and the high number of neurons, the processing time was be very long and after 500 iterations the network was capable of recognizing only 2 classes. Therefore, the training has also been performed through an algorithm that, for a specified number of iterations, selects randomly a sample for each class and presents them to the neural network.
- A "simplified" legend with 15 classes. The classes have been grouped in categories that have similar characteristics. We have tried to chose classes that are most important for hydrological modeling, but at the same time we have considered the spectral characteristics that can discriminate between one class and the other. The different types of vegetation (pine trees, firs, oaks, chestnuts and so on), have been grouped in the two classes "Evergreen" and "Deciduous" Forest; classes like "Urban Residential", "Urban Industrial", "Urban Infrastructures" are grouped in "Urban Areas"; classes like "Frutescent Mediterranean Maquis" and "Arboreal Mediterranean Maquis" are grouped in a single class.

4 PRELIMINARY RESULTS AND DISCUSSION

The Landsat Thematic Mapper images have been classified with the techniques described above.

Cluster Analysis and Maximum Likelihood Classification have been performed with ERDAS IMAGINE® 8.3.1 software.

We have then calculated the number of pixels in the reference data that were correctly classified with the different methods. The results obtained are summarized in Table 1 (the 69 classes have been grouped in major land cover types).

<table>
<tr><th>Land cover class</th><th>ISODATA Clustering</th><th>Maximum Likelihood</th></tr>
<tr><td>Cultivated land</td><td>76.1%</td><td>46.8%</td></tr>
<tr><td>Urban Areas</td><td>40.2%</td><td>80.5%</td></tr>
<tr><td>Evergreen Forest</td><td>59.6%</td><td rowspan="2">74.4% *</td></tr>
<tr><td>Deciduous Forest</td><td>39.9%</td></tr>
<tr><td>Water</td><td>40.7%</td><td>66.7%</td></tr>
</table>

Table 1: Percentages of pixels of the training sample correctly classified with Unsupervised and Supervised Multispectral Classification (*Note: in Supervised classification evergreen and deciduous forest are grouped in the same class)

To implement the Neural Networks classifiers, we have used MATLAB® Neural Network toolbox version 3.0.

In the case of the Neural Network with 69 classes, if we use the entire sample as training set, as previously said the training is very long and the Network is not capable of discriminating between classes when the training is stopped after an acceptable number of iterations. A very high number of iterations would be required, and still there would probably be some confusion between classes that have very similar spectral response.

If we use randomly chosen inputs for training, the Network is faster in classifying Land Cover Types but the results are not satisfactory for classes that have a high variability, i.e. that actually include different kind of surfaces (cultivated land). In fact, using random samples and not the entire dataset it is possible that some kind of surfaces are not included in the training and then are not recognized. In this case, the results can be improved by executing additional training presenting only inputs from selected classes (those which have shown worst performance). Anyway this additional training, while improving the accuracy for some classes, could make it worse for others. Therefore, this method could be applied to all classes that present an unsatisfactory classification, but has to be supervised until acceptable results for all the land cover classes are obtained.

The Neural Network with 15 output classes still needs many iterations of training due to the high number of samples (all data are presented to the network), but after 25000 epochs it is already possible to recognize the major classes (cultivated Land, urban areas, Deciduous/Evergreen Forest, Water). Other classes that have much smaller number of pixels are not recognized. The class which has the majority of pixels (Cultivated Land) is predominant and classes like "Cultivated Land with trees", "Oliveyard", "Vineyard", "Pastures" turn out to be classified as "Cultivated Land".

Tables 2 and 3 shows the error matrix after 25000 iterations and the percentage of pixels in the training sample that are correctly classified for the major Land cover classes. Element *(i ,j)* in the error matrix represent the number of pixels classified as class i by the network that actually are class j in the training sample. Thus, the diagonal elements in the matrix turn out to be the number of pixels correctly classified. The mean overall accuracy (sum of the diagonal elements divided by the total number of pixels) is 48.3%.

It should be noted that these results are relative to the training dataset, so they give information only about the classification performance on the training data. Once the results are considered acceptable for these data, their application on the full image should be verified on another set of ground truth points, different from the one used for training. At the same time, the error matrix is useful for identifying “critical” aspects of classification procedures, for instance which classes are spectrally separable and which are not.

NEURAL NETS CLASSIFICATION / GROUND TRUTH	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	660	252	38	45	99	36	57	43	58	9	21	38	4	28	13
2	911	4540	418	669	755	217	540	201	96	171	115	725	65	144	79
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	40	93	41	41	272	4	81	88	21	31	35	48	-	5	2
6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	70	254	54	65	240	4	2865	853	313	109	178	260	17	85	17
8	51	96	16	25	113	7	391	1029	59	39	253	43	27	14	20
9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	-	2	-	1	-	-	3	1	-	1	1	9	-	3	-
13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	1	-	-	-	-	-	-	1	-	-	-	-	-	-	-
15	5	8	-	1	-	-	3	-	-	1	1	1	5	11	57
TOT	**1738**	**5245**	**567**	**847**	**1479**	**268**	**3940**	**2216**	**547**	**361**	**604**	**1124**	**118**	**290**	**188**

Table 2: Error Matrix after 25000 iterations. (Neural Network classifier with 15 output classes). Training sample = 19532 pixels
1.Urban Areas 2.Cultivated Land 3.Cultivated Land with trees 4.Vineyard 5.Oliveyard 6.Nurseries, greenhouses 7.Deciduous Forest 8.Evergreen Forest 9. Bare/Degraded Land (landslides/fires) 10.Brushland 11.Mediterranean maquis 12.Patures and grassland 13.Marsh Vegetation 14.Quarries and landfills 15.Water

Land Cover Class	***% of pixels correctly classified***
Cultivated Land	86.6%
Urban Areas	38.0%
Deciduous Forest	72.7%
Evergreen Forest	46.4%
Water	30.3%

Table 3: Percentages of pixels of the training sample correctly classified with Neural Network classifier after 25000 iterations.

These preliminary results confirm the potential of Neural Networks as satellite images classifiers; although the number of iterations required for classes that have a small number of samples may be very high. With a lower number of epochs, classes with more sample pixels are predominant.

The performance is better if the classes are grouped in more general categories that have similar spectral characteristics; in this case in fact the training is faster and the confusion between classes is reduced.

In some cases, unsatisfactory results can be due to errors in the "ground truth" data and not to the Neural Network. For example, the ground truth data can be incorrectly geolocated: a comparison with drainage network shows that some pixels classified as "water" are actually misplaced.

Further work will focus on the evaluation of results after a higher number of iterations, and how the classification improves with training.

Also, the effect of training dataset on the network performance could be investigated: for example a quality assessment of ground truth data to eliminate the pixels that give higher error or are not correctly geolocated.

Acknowledgments

The authors would like to thank GEOS S.r.l for providing the ground truth data, and Prof. Ignazio BECCHI, for his help and suggestions.

REFERENCES

Bischof H., Leonardis A. Finding optimal Neural Networks for Land Use Classification, *IEEE Transactions on Geoscience and Remote Sensing*, 1998, **36**(1), 337-341.

Bischof H., Schneider W., Pinz A.J., Multispectral Classification of Landsat images using Neural Networks, *IEEE Transactions on Geoscience and Remote Sensing*, 1992, **30**(3), 482-490.

Callan R., The essence of Neural Networks, Prentice Hall Europe, 1999.

Castellani L., Becchi I., Castelli F., Rainfall frequency and seasonality identification through artificial neural networks, *Meccanica,* 1996, **31**, 117:127.

Fardanesh M.T., Ersoy O.K., Classification Accuracy Improvement of Neural Network Classifiers by using Unlabeled Data, *IEEE Transactions on Geoscience and Remote Sensin*g, 1998, **36** (3), 1020-1025

Lillesand T.M., Kiefer R.W., Remote Sensing and Image Interpretation, II ed., John Wiley & Sons, 1987.

Rees W.G., Physical principles of remote sensing, Cambridge Univ. Press, 1990

Regione Toscana - Giunta Regionale, Verso il sistema informativo territoriale della Regione Toscana:, Tuscany Region Publication, 1996.

Tou J., Gonzales R.C, Pattern Recognition Principles, Addison Wesley Publishing Company, 1974.

Yoshida T., Omatu S., Neural network approach to land cover mapping, *IEEE Transactions on Geoscience and Remote Sensing*, 1994, **32**(5), 1103-1109.