Land-cover classification in remote-sensing images using structured neural networks

F. Roli and S. B. Serpico

Dept. of Biophysical and Electronic Eng. – University of Genoa
Via All' Opera Pia 11A, 16145, Genoa, ITALY
Ph. +39-10-3532672 Fax +39-10-3532134 e-mail: roli@dibe.unige.it

ABSTRACT

This paper proposes the application of structured neural networks to land-cover classification in remote-sensing images (in particular, multisensor images are considered). Purpose of our approach is to give a criterion for network architecture definition and to allow the interpretation of the "network behaviour". The first result aims to avoid a cumbersome trial-and-error process; the latter one can be used to obtain information about the relevance of sensors and related bands to land-cover classification. First of all, the architecture of structured networks is tailored to a multisensor classification problem. Then, they are trained and transformed into "simplified networks" which allow one to evaluate the relevance of sensors and related bands. Experimental results on a multisensor data set related to an agricultural area are reported. An additional experiment on a multispectral image of a forest area is also briefly described. Comparisons with the Bayesian classifier confirm the effectiveness of our approach.

1. INTRODUCTION

One of the primary uses of remotely-sensed images is in land-cover classification for applications such as agricultural and forest monitoring (Townshend, 1992). Notwithstanding the considerable potential demonstrated by techniques for land-cover classification, improvements are necessary to achieve satisfactory performances. The use of multisensor images seems to be a promising direction to obtain such performances (Townshend, 1992). Consequently, many efforts have been devoted to the development of new approaches for the classification of multisensor images (Luo and Gray, 1989). The interest of applying Artificial Neural Networks (ANNs) has recently been documented in various works. Benediktsson *et al.* (1990) compared two approaches to multisource classification based on neural networks (supervised learning by the delta rule and by the back-propagation algorithm) with classical statistical methods. A new neural network architecture, based on concepts from the Consensus theory, was proposed and experimented on multisource data by Benediktsson *et al.*, 1991. Bischof *et al.* (1992) investigated the classification of multispectral images by a one-hidden layer back-propagation ANN; they also performed texture information extraction and post-classification smoothing by neural processing.

Neural network approaches provide, in fact, important advantages: no need for a priori knowledge on statistical distribution of data, intrinsic parallelism, fast classification time, fault tolerance. On the other hand, well-known problems related to the use of ANNs are to be faced: no general criteria for defining a suitable architecture of the network, difficult interpretation of the "network behaviour" (the so-called "opacity problem"), dependence of classification performances on various factors (initial weights, choice of the training set, training parameters, etc.). So far, such problems have only been partially considered in the literature of ANNs for remote-sensing data classification.

This paper proposes the application of structured neural networks to land-cover classification in multisensor remote-sensing images. The purpose of this approach is to exploit ANNs' advantage while solving, in the context of the considered application, the problems of "architecture definition" and of "opacity".

In Section 2, we propose a class of structured neural networks and a criterion to define a specific architecture for a given multisensor classification problem. The transformations applied to simplify the representation of our ANNs in order to interpret the "network behaviour" are dealt with in Section 3. Experimental results are reported and discussed in Section 4. Conclusions are given in Section 5.

2. STRUCTURED NEURAL NETWORKS FOR CLASSIFICATION OF MULTISENSOR REMOTE-SENSING IMAGES

The proposed approach addresses the problem of supervised classification of multisensor remote-sensing images, assuming that the preceding image processing steps (correction, registration, pattern and feature extraction) have already been carried out by whatsoever techniques. In particular, our approach is based on multilayer feedforward networks (commonly called "multilayer perceptrons", Hertz *et al.*, 1991). A novel aspect concerns "architecture definition". It is aimed to make possible the interpretation of the "network behaviour" (Sections 2.1 and 3.2) and to take explicitly into account the peculiarities of multisensor data (Section 2.2).

2.1 Adopting an architecture based on tree-like networks

In a fully-connected layered ANN, all neurons of a layer contribute to the input to every neuron of the successive layer (Hertz *et al.*, 1991). The resulting "distributed" information processing hinders the interpretation of the "network behaviour". Therefore, we propose to adopt architectures for which the output of each neuron is fed as input to just one neuron of the next layer, that is, tree-like networks (TLNs). In this way, well-separated neuron contributions are used (details on the architecture of our TLNs are given in the following Section).

The global TLN-based architecture we propose is depicted in **Figure 1**. A multiple-input single-output tree-like network is devoted to each class of data; the outputs of all the TLNs are then compared by a decision block that makes the final decision about classification. In our approach, we train each TLN separately so that its output provides an estimate of the posterior probability of the related class (Makhoul, 1991). According to the Bayes rule, we want to assign samples to the estimated most probable class; that is, we want to select the TLN with the maximum output. Therefore, we have to adopt a "winner-takes-all" decision block **(Figure 1)**.

To complete the interpretation of TLN behaviour we must also evaluate the importance of each connection with respect to the network output; this aspect will be considered later on (Section 3).

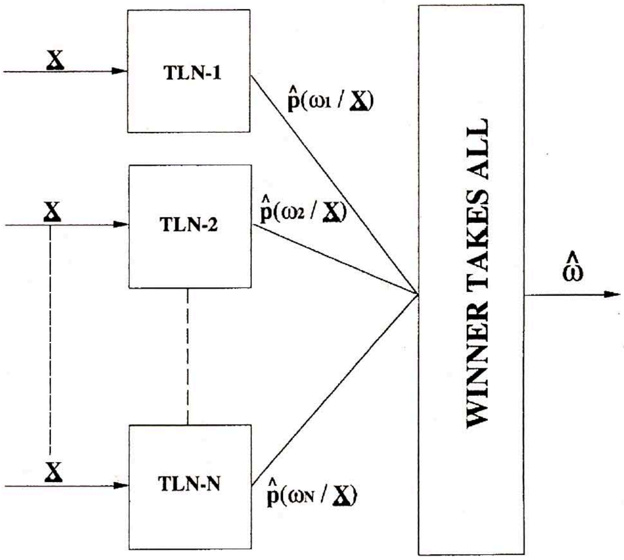


Figure 1 - The Global TLN-based Architecture

2.2 Tailoring the architecture to a multisensor classification problem

According to our method, the architecture of the class-related TLNs are all identical **(Figure 1)**; therefore, we will refer in the following just to a unique TLN architecture.

The operation of each TLN may be interpreted as a *check on a criterion* that can be hierarchically decomposed into subcriteria (a similar interpretation has been proposed also in (Krishnapuram and Lee, 1992)). In particular, each layer of neurons in a TLN corresponds to a level of decomposition of criteria into subcriteria. We propose that, as first decomposition level, the *global criterion* checked by the output neuron be decomposed into *sensor-related subcriteria* (one subcriterion per each available sensor). Then, as second decomposition level, each sensor-related subcriterion be decomposed directly into *elementary subcriteria*, defined on the inputs derived from the corresponding sensor **(Figure 2)**.

Such an interpretation of a TLN notably helps the definition of a network architecture tailored to a given multisensor classification problem. In fact, it defines a two-hidden layer architecture of the type in **Figure 2**. Such an architecture is completely defined except for the number of neurons of the first hidden layer, which depends on the complexity of the elementary subcriteria we adopt.

The TLN-based architecture obtained in this way explicitly take into account the multisensor nature of the problem considered and allows one to easily configure the network architecture. In particular, the information derived from different sensors is processed separately inside subnets, then combined only at the level of the output neuron.

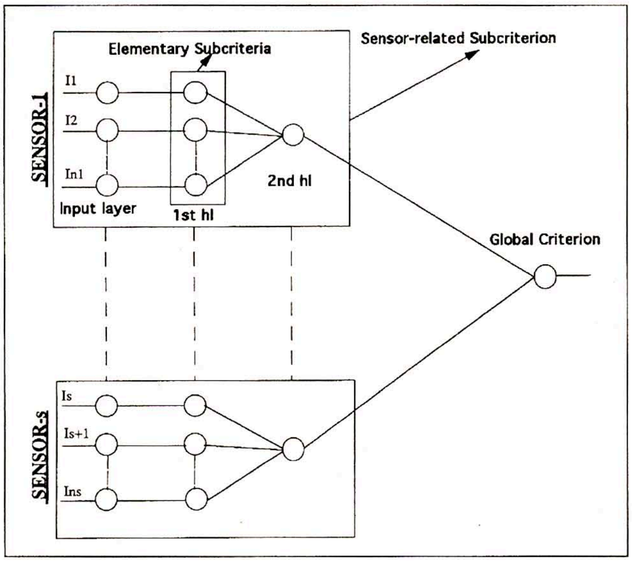


Figure 2 - Tree-like Network (TLN) tailored to a multisensor classification problem (hl=hidden layer)

3. SIMPLIFICATION AND INTERPRETATION OF A TREE-LIKE NETWORK

In this Section, we propose a technique to generate a "simplified representation" of our TLNs (Section 3.1) which allows one to interpret the network behaviour (Section 3.2).

3.1 Simplification

Let us now consider why the weight of a connection does not correspond directly to its importance. First, for all possible values of the inputs to a TLN, the output of some neurons may exhibit variations in a small subrange of their full output dynamics; the output connections of these neurons provide reduced contributions to the next layer's neurons, with respect to a full usage of output dynamics. Secondly, the importance of a connection depends also on the bias of the neuron the connection enters and on the weights of the remaining connections entering the same neuron; the presence of both positive and negative weights makes interpretation still more difficult. Finally, dealing with complicated mathematical functions (*e.g.*, the "logistics function"), used to describe the neurons' response, is not intuitive for a human being.

In order to overcome this kind of difficulties, we propose to define a TLN to be used for classification, to train it and then to transform the representation of such a network so that it may exhibit the following properties:

- weights are all positive;
- the output of all neurons vary over the whole range of their output dynamics;
- weights and bias entering each neuron are normalised to the sum of the weights themselves;
- the neurons' response is approximated by a linear function with saturation.

TLNs with the above properties are obtained by four progressive transformations:

i) the first step is the transformation of a generic TLN into an equivalent one with positive weights (TLN^+). We do not apply this procedure to the weights of the connections between the neurons of the input layer and those of the first hidden layer, as our method does not require that these weights be positive (Serpico *et al.*, 1992).

ii) the second step is the transformation of a TLN^+ into an equivalent network in which the output of every unit of hidden layers is expanded to the whole range of the neuron non linearity output ([0,1], in our case).

iii) normalization: it is applied to all weights and biases of the "expanded" TLN. The result is an equivalent network in which the weights of all the connections entering each unit are normalized by a multiplicative factor such that their sum be equal to a prefixed positive value N (*e.g.*, N=1000). Each bias is normalized by using the same multiplicative factor as for the input weights to its unit.

The value of each normalized connection is called "Voting Power" of the neuron which generates such a connection (the reason for such a name will be clarified in the next Section).

iv) Finally, to simplify the neuron non-linear function of the normalized network, we approximate it by a function linear with saturation to its minimum and maximum values (*i.e.*, 0 and 1, respectively). In particular, we utilize the straight line that is tangent to the neuron non-linear function at the point where this function is equal to 0.5. In order to characterize the resulting piecewise linear function **(Figure 3b)**, we use two parameters: the "Voting Threshold" (VT) and the "Delta Votes" (ΔV).

The meaning of such parameters will be clarified in the next Section. VT is defined as the value that makes the piecewise linear function equal to 0.5; ΔV is defined as the difference between VT and the value corresponding to the extremes of the sloping part of the linear function **(Figure 3b)**.

The algorithms to perform the above transformations are given in (Serpico *et al.*, 1992). The resulting networks constitute a "simplified representation" of the original TLNs. They are called "Piecewise Linear Tree-Like Networks" (PLTNs). An example of PLTN is given in **Figure 3**. Such networks are equivalent to the original TLNs, except for the piecewise linear approximation of the logistics function.

3.2 Interpretation

According to the above definition of PLTNs, the behaviour of original TLNs can be easily interpreted. In particular, the hierarchical arrangement of criteria described in Section 2.2 can be interpreted as a hierarchical arrangement of *committees* (Roli *et al.*, 1993).

The interpretation of input units is obvious: they provide a coded representation of the data to be classified.

In order to interpret the role of the remaining units in the net, we recall that inside a PLTN all weights from the first hidden layer (hl) on are positive numbers and all neuron nonlinearities are monotone increasing functions. Therefore, first hl units provide positive contributions to neurons of the second hl, which are propagated again as positive contributions to the output unit. In particular, the outputs of the first hl units are multiplied by the normalized weights (w_{norm}) of the output connections by which they are propagated, hence their contribution to second hl units vary in the range $[0,w_{norm}]$. Recalling that a high output value of a network suggests the hypothesis that a sample may be assigned to a given class, we say that a unit may assign a number of votes in the range $[0,w_{norm}]$ to this hypothesis. Hence, each unit of the first hl is regarded as a member of a *sensor-related committee*, and the quantity wnorm (*i.e.* the normalized weights) is called the "Voting Power" (VP) of that member (**Figure 3**). The sum of the voting powers inside a committee is normalized to N (*e.g.*, N=1000).

The second hl units are interpreted as "vote-taking units" (VTUs) of the above sensor-related committees. Therefore their piecewise linear responses give "majority rules": to have a positive judgement (*i.e.*, an output greater than or equal to 0.5) at least VT_j votes are required; an increase or decrease of at least ΔV_j votes causes the judgement reach the value 1 or 0, respectively. Consequently, VT_j is called the "Voting Threshold" and ΔV_j the "Delta Votes" of the j-th sensor-related committee.

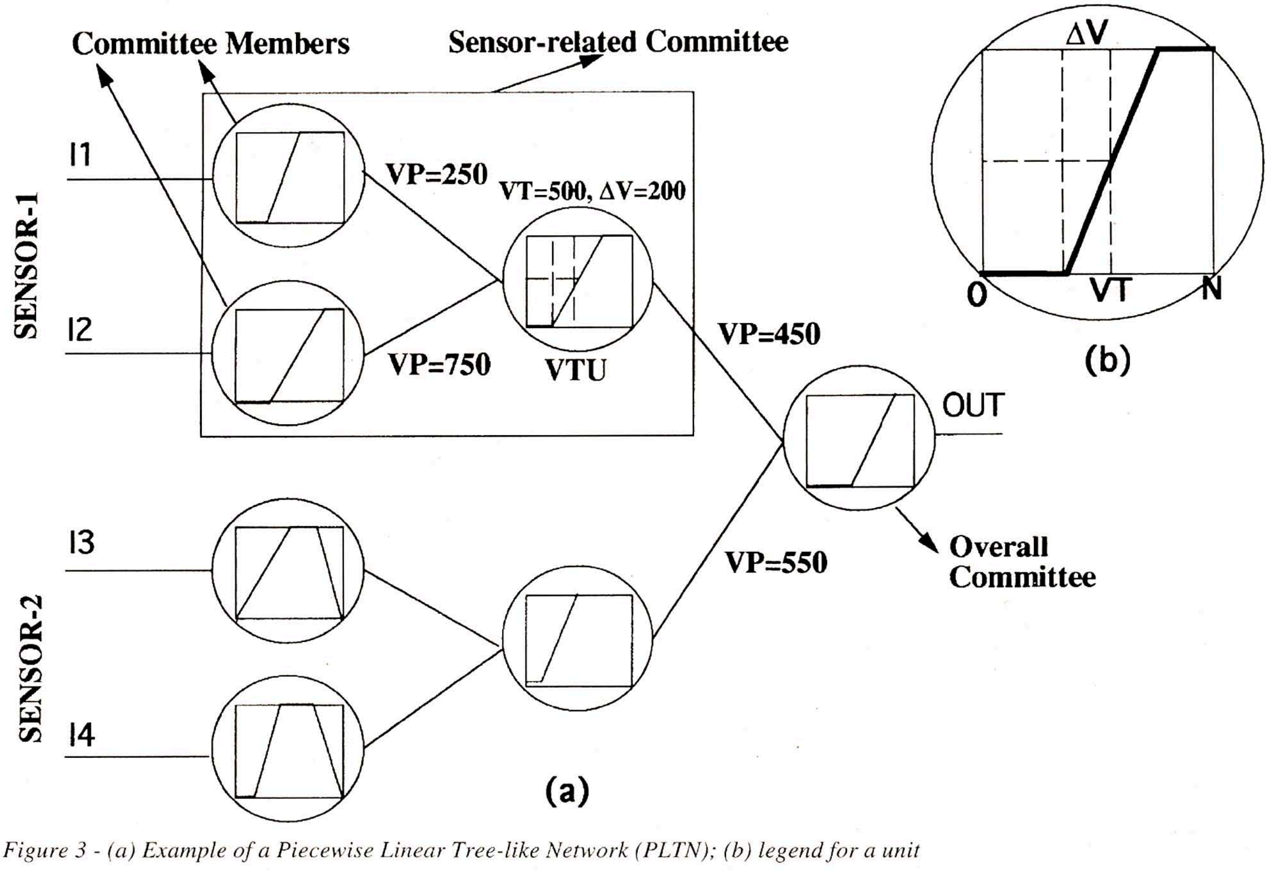


Figure 3 - (a) Example of a Piecewise Linear Tree-like Network (PLTN); (b) legend for a unit

At the same time, the second hl units are the members of a further committee, that is, the "overall committee". They use their judgement to decide how many of their available votes (VP) assign to the current classification hypothesis. The role of the output unit is that of VTU of the overall committee.

The task of all of the above committees is to judge if a sample belongs or not to a given class. Inside all of them, the sum of the voting powers of the component members is equal to a prefixed number N.

The voting power of a member of a sensor-related committee tells how much important is a elementary criteria, therefore, also how much important is the related input (*e.g.*, a sensor channel). The voting power of a member of the overall committee is proportional to the importance of the related sensor. In both cases, the importance is related to the assignment of samples to a given class.

Each committee corresponds to a criterion or subcriterion defined in Section 2.2, and the non-linear functions used to aggregate subcriteria into criteria correspond to the linear majority rules of the corresponding committees.

4. EXPERIMENTAL RESULTS

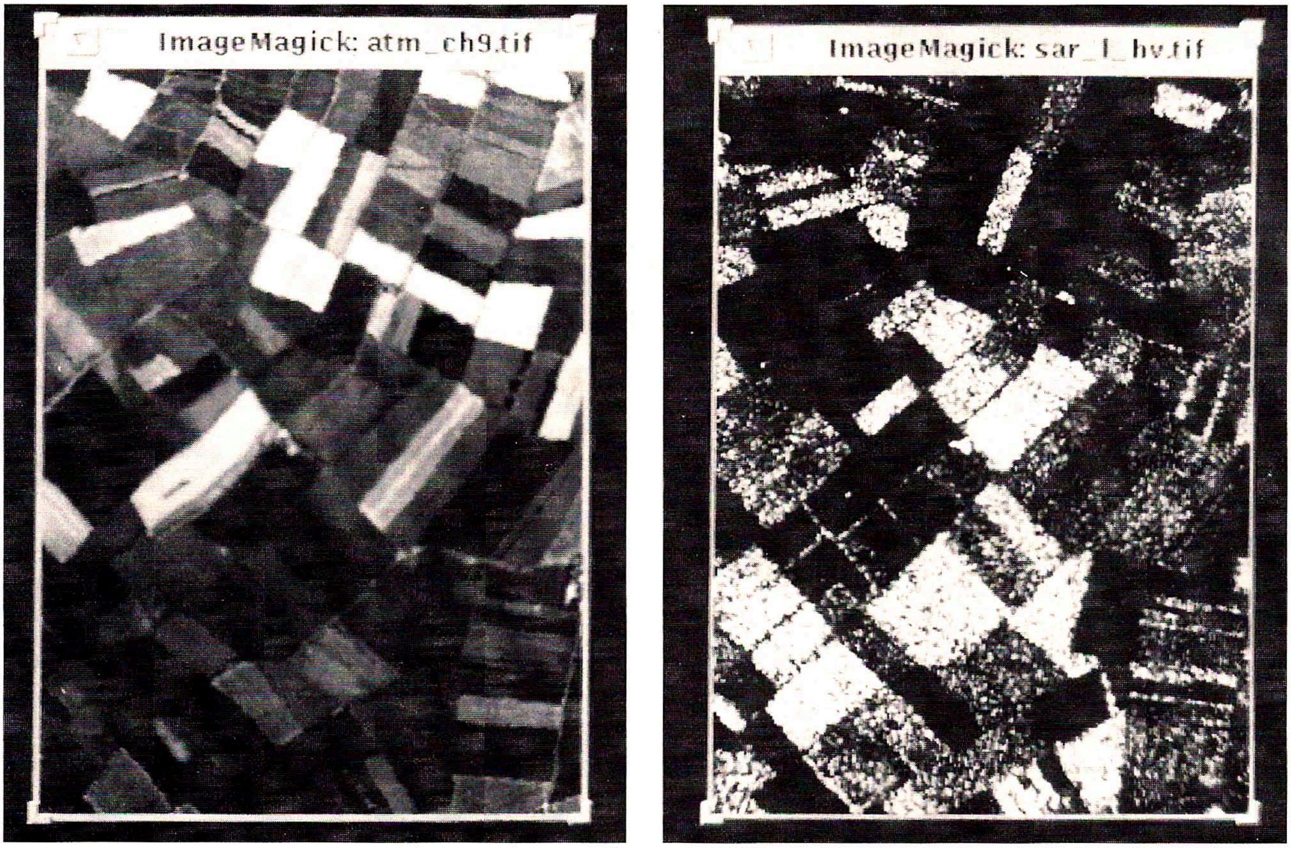
Experiments were carried out on various data sets to validate the effectiveness of our approach for land-cover classification. In the following, we will focus on an agricultural application concerning crop discrimination. An additional experiment on a multispectral image of a forest area is also briefly described. For other experiments, we refer the reader to (Serpico *et al.*, 1993, Roli *et al.*, 1993).

4.1 Agricultural application

The data set consists of multisensor remote-sensing images acquired by two sensors installed on aircraft: a Daedalus 1268 Airborne Thematic Mapper (ATM) scanner, and a PLC band, fully polarimetric, NASA/JPL imaging radar system. The former is a passive sensor working in the visible and infrared portion of the electromagnetic spectrum; the latter is an active sensor providing synthetic aperture radar (SAR) images. The flights took place in July and August 1989, respectively.

Sub images of 250 x 350 pixels, related to the agricultural area of the Feltwell village (U.K.), were selected for classification experiments **(Figure 4)**. Images were regis-

(a)



(b)

Figure 4 - (a) ATM image; (b) SAR image

tered to an average accuracy of about 1 pixel by using the radar image as reference and scaling and registering the ATM image on it. Based on ground truth, we selected pixels from the five numerically most representative agricultural classes (sugar beet, stubble, bare soil, potatoes, carrots). For each pixel, a feature vector was computed by utilizing the intensity values in six ATM bands and 9 features extracted from radar images (mean values were extracted from 9 radar channels by using a 9 x 9 window surrounding each pixel; we selected P, L, and C bands with polarizations HH, VV, and HV). For input coding, we normalized the above fifteen features in the range [0,1].

4.1.1 Classification Performances

We used five single-output TLNs, one for each data class **(Figure 5)**. A training set was generated by randomly selecting 50 % of the 8836 pixels. The remaining pixel were used as test patterns.

Each TLN was separately trained in order to classify training set pixels as belonging or not to the related data class. The error backpropagation training procedure was used (training by epoch). As convergence criterion we adopted a threshold value on the " global mean square error". The five TLNs obtained after the training procedure were connected to a winner-takes-all block (as in **Figure 1**) and used to classify test patterns. **Table 1** shows the classification results on training and test data. The classification results provided by a Bayesian classifier (assuming a normal distribution for each data class) are also reported.

4.1.2 Interpretation of the network behaviour

For instance, let us consider how the behaviour of the TLN devoted to the class "Bare Soil" can be interpreted according to the proposed method. To interpret its behaviour, we consider the "simplified representation" in **Figure 6**, that is, the related PLTN.

Such a graphical representation gives a synthetic view of the neural network behaviour. For example, we can notice that intensity in ATM bands is a little more important than the intensity in radar channels. In fact, the weights associated to ATM and SAR subnets are 585 and 415, respectively. For the other data classes the importance ("significance parameter") of the considered imaging sensors (SAR and ATM) is shown in **Table 2**. The mean values indicate that both sensors are important to solve the given multisensor classification problem.

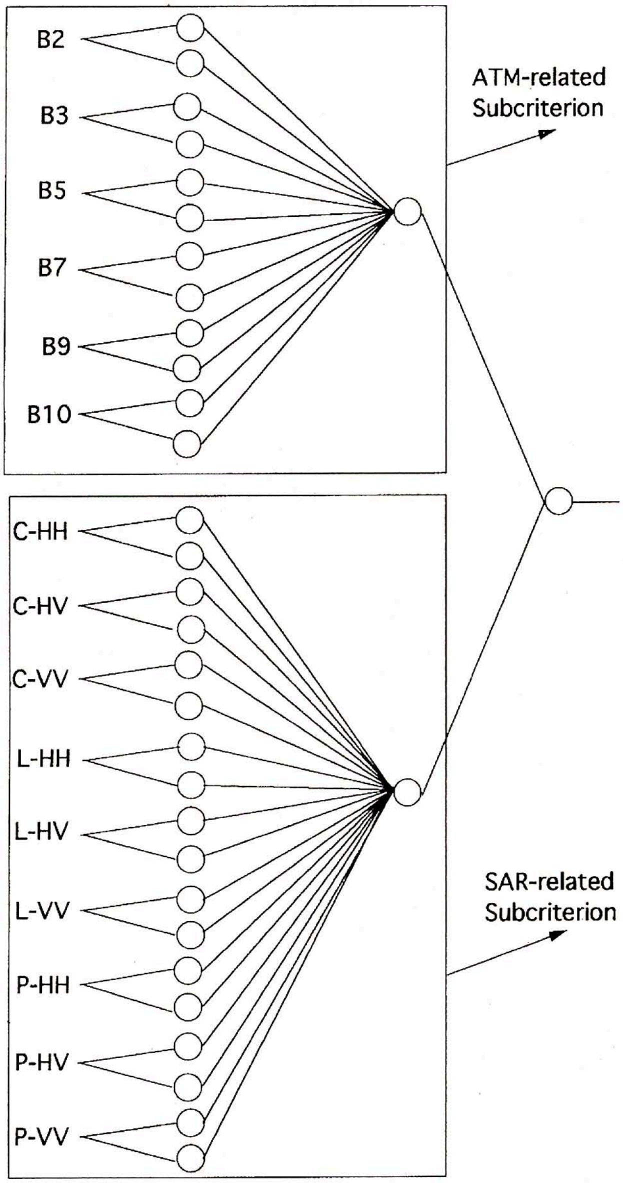
We can consider the Voting Powers and the Voting Thresholds related to ATM and SAR committees to evaluate the importance of the related bands. With regard to the ATM sensor, we can say that the 3rd band seems to be negligible. Its Voting Power is much smaller than the Voting Powers related to the other bands. Band C of the SAR sensor seems to be the most important. In fact, the sum of the voting powers related to band C channels can exceed the Voting Threshold (VT=590) of SAR committee. On the contrary, channels related to band P seem to be negligible.

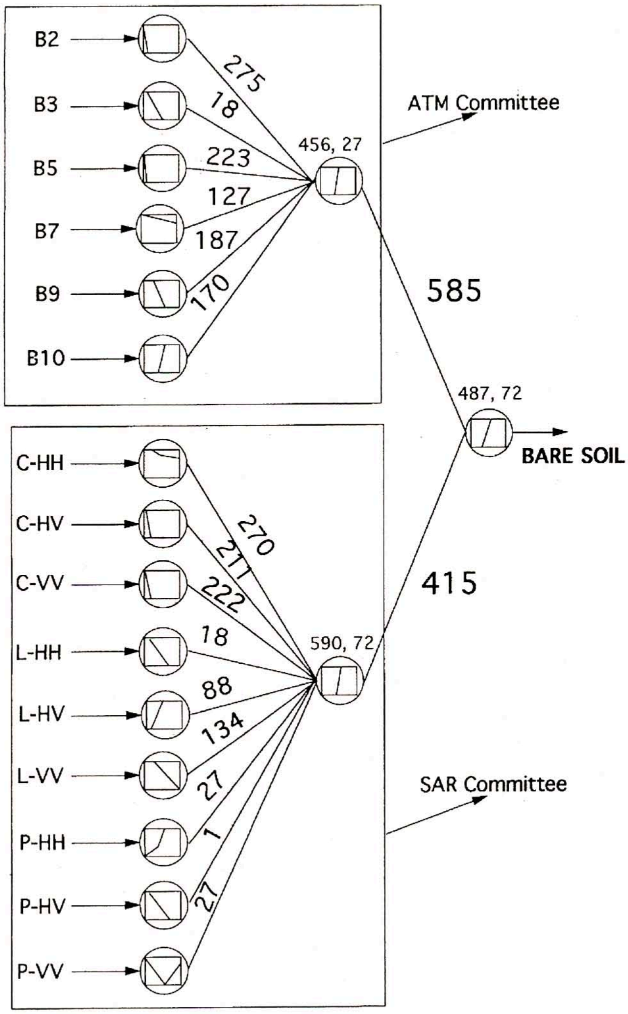
4.2 Forest application

This data set was obtained by using the above described ATM image, only. It consists of 100 regions obtained by segmenting an area of 255 x 255 pixels including the Thetford forest (close to the Feltwell village in U.K.). We selected the numerically most representative class, that is, according to the ground truth, "deciduous forest". The class label of each region indicates belonging or not to deciduous forest. The average value in five ATM bands were selected as spectral features. In addition, we considered a set of three geometrical features computed for each region: the region size (*i.e.*, the number of pixels belonging to a region), and two shape parameters. These last parameters are computed by first extracting the Minimum Bounding Rectangle (MBR) for each region (that is the smallest circumscribed rectangle), then computing the ratio between the longest and the shortest side of the MBR ("elong"), and the ratio between the region size and its MBR size ("fit").

Due to the small number of samples, we have adopted the *k-fold cross-validation* process (with k=4) to estimate classification performances (Stone, 1974). Consequently, we subdivided the dataset in four subsets, then we used three of them as training set and the remaining one as test set. This was done in all of the four possible combinations and the mean value of classification performances was computed.

The TLN shown in **Figure 7a** was trained to classify pixels as belonging or not to the "deciduous forest" data class. Such a network is constituted by two sensor-related subcriteria. Geometrical features ("region size", "elong", and "fit") were regarded as information derived from a "virtual" sensor. Classification performances estimated by the k-fold cross-validation process were good (an error rate equal to 1.74 %). Network behaviour was interpreted by the related PLTN **(Figure 7b)**. ATM sen-





↑ *Figure 6 - PLTN for the "Bare Soil" data class (the related TLN has the architecture depicted in Figure 5)*

← *Figure 5 - Architecture of TLNs used to classify multisensor images in Figure 4 (one TLN for each data class)*

Table 1 - Classification performances of the proposed neural networks in comparisons with a Bayesian classifier

	STRUCTURED ANNs		BAYESIAN CLASSIFIER	
DATA CLASS	% ERROR RATE (TRAINING)	% ERROR RATE (TEST)	% ERROR RATE (TRAINING)	% ERROR RATE (TEST)
SUGR BEET	2.57	3.66	1.69	1.96
STRUBBLE	6.78	8.5	12.35	11.23
BARE SOIL	12.65	12.69	9.04	9.67
POTATOES	2.75	3.48	2.01	3.48
CARROTS	3.99	4.15	5.77	8.15
GLOBAL % ERROR RATE	4.52	5.45	5.32	5.88

sor seems to be more important than the "virtual" sensor related to geometrical features(693 votes as compared with 307 votes). In particular, two geometrical features ("region size" and "fit") seem to be negligible on the basis of their Voting Powers (73 and 120). This last conclusion was confirmed by removing such features and testing classification performances. After retraining, very similar performances as with the complete set of features were obtained.

Table 2 - Evaluation of the importance of the considered imaging sensors

DATA CLASS	SIGNIFICANCE PARAMETER ATM SENSOR	SAR SENSOR
SUGAR BEET	0.518	0.482
STUBBLE	0.382	0.618
BARE SOIL	0.585	0.415
POTATOES	0.529	0.471
CARROTS	0.540	0.460
MEAN VALUE	0.510	0.490

5. CONCLUSIONS

In this paper, we have presented a novel approach to land-cover classification in multisensor remote-sensing images. Such an approach is based on structured neural networks which allows one to exploit ANNs' advantages while solving, in the context of the considered application, the problems of "architecture definition" and of the interpretation of the "network behaviour". From the viewpoint of the land-cover classification, the proposed method provides important advantages. It constitutes a simple "data fusion" technique to integrate the information extracted by multisensor images. In addition, it allows one to evaluate the importance of different sensors and of their bands. With regard to classification performances, comparisons with the Bayesian classifier confirmed the effectiveness of the proposed approach.

This research work could be expanded in many directions. In particular, the proposed approach could be easily applied to multisource classification problems (Benediktsson *et al.*, 1990), by devising an appropriate input coding for non-numerical sources.

ACKNOWLEDGEMENTS

This work was partially supported by ASI (Italian Space Agency) and by the cooperative research project MEDALUS II (Mediterranean Desertification and Land Use) which was funded by EEC (Environment Programme, Contract CT-EV-5V-0166). The authors wish to thank HTS – Hunting Technical Services (UK) for supplying ATM images and Profs. Gianni Vernazza and Gianni Jacucci for their helpful suggestions and comments.

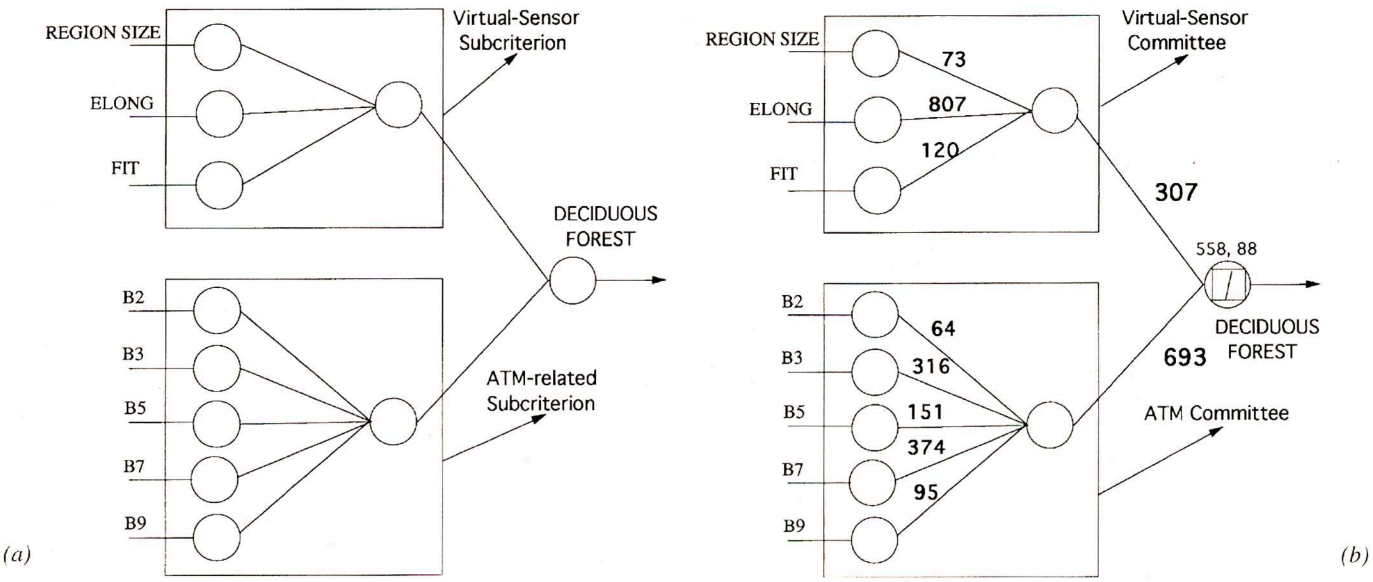


Figure 7 - (a) TLN used to classify pixels belonging to "deciduous forest" data class; (b) The related PLTN

REFERENCES

Benediktsson J.A., Swain P.H., Ersoy O.K., "Neural Network Approaches versus Statistical Methods in Classification of Multisource Remote-Sensing Data," IEEE Trans. on Geoscience and Remote Sensing, Vol. 28, No. 4, pp.540-552, July 1990.

Benediktsson J.A., Ersoy O.K. & Swain P.H.,"A Consensual Neural Network", Proc. of the 1991 Int. Geoscience and Remote Sensing Symp., IGARSS 91, Espoo, Finland, IEEE Press, June 3-6, 1991, pp. 2219-2222.

Bischof H., Schneider W., & Pinz A.J., "Multispectral Classification of Landsat-Images using neural networks", IEEE Trans. on Geoscience and Remote Sensing, Vol. 30, No. 3, May 1992, pp. 482-490.

Hertz J., Krogh A. & Palmer R.G., Introduction to the Theory of Neural Computation, Addison Wesley Pub. Co., The Advance Book Program, 1991.

Krishnapuram R. & Lee J., "Fuzzy-Set-Based Hierarchical Networks for Information Fusion in Computer Vision", Neural Networks, Vol. 5, No. 2, 1992, pp. 335-350.

Luo R.C. and Gray M., "Multisensor integration and fusion in intelligent systems", IEEE Trans. on Systems, Man, and Cyb., Vol. 19, No. 5, pp. 901-931, September/October 1989.

Makhoul J., "Pattern Recognition Properties of Neural Networks", Proc. of the 1991 IEEE Workshop on Neural Networks for Signal Processing, IEEE Press, 1991, pp. 173-187.

Roli F., Serpico S.B. & Vernazza G., "Multisensor image classification by structured neural networks", Proc. of the 1993 IEEE Workshop on Neural Networks for Signal Processing, IEEE Press, Baltimore, MD, USA, September 7-9, 1993, pp. 311-320.

Serpico S.B., Roli F. & Vernazza G., "Classification of multisensor remote-sensing images by structured neural networks", Technical Report, D.I.B.E. – Univ. of Genoa, NN-12-92, Dec. 1992.

Serpico S.B., Roli F., Pellegretti P. & Vernazza G., "Structured neural networks for the classification of multisensor remote-sensing images", Proc. of the 1993 Int. Geoscience and Remote Sensing Symposium (IGARSS'93), Tokio, Japan, August 18-21, 1993, pp. 907-909.

Stone M., "Cross validation choice and assessment of statistical predictions", Journal of the Royal Statistical Society, 36, 1974, pp. 111-147.

Townshend J.R.G., "Land Cover", Int. Journal of Remote Sensing, Vol. 13., Nos. 6 and 7, pp. 1319-1328, 1992.