Autoassociative Neural Networks for Features Reduction of Hyper-Spectral Data

Fabio Del Frate*, Giorgio Antonino Licciardi*, Riccardo Duca**

**Department of Computer Science, Systems and Production , Tor Vergata University, Rome, ITALY*

(delfrate@disp.uniroma2.it - licciardi@disp.uniroma2.it)

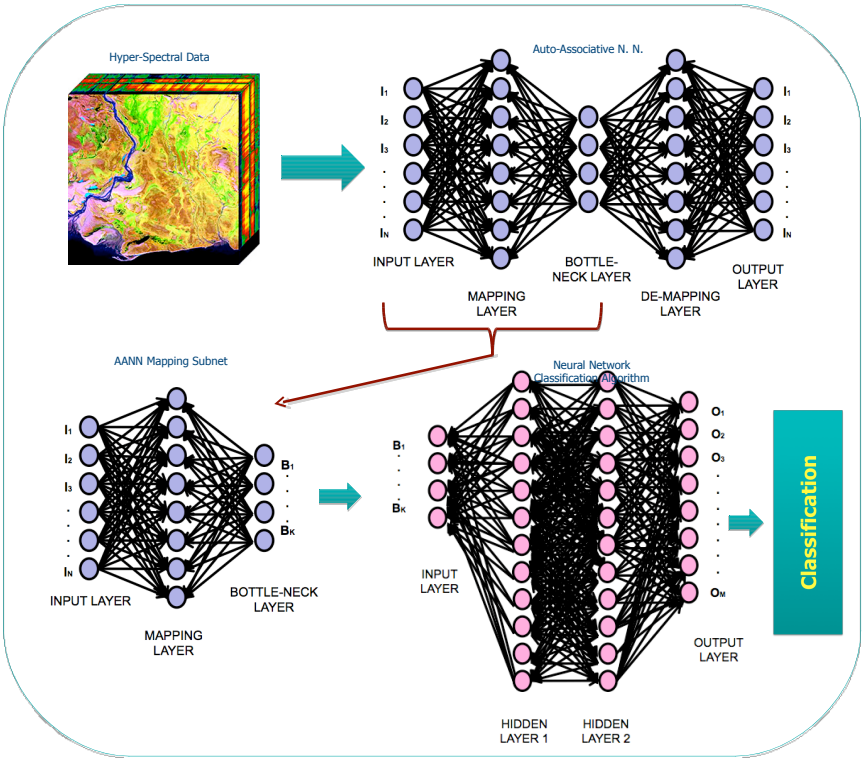
***ESA-ESRIN, Via Galileo Galilei 00044 Frascati, Italy*

Abstract - The potential of neural networks has been applied to hyper-spectral data, and exploited either for classification purposes and for feature extraction and dimensionality reduction. For this task a topology named auto-associative neural network has been used. The processing scheme uses a neural network architecture consisting of two stages: the first stage reduces the dimension of the input vector, while the second stage performs the mapping from reduced input into land cover classification.

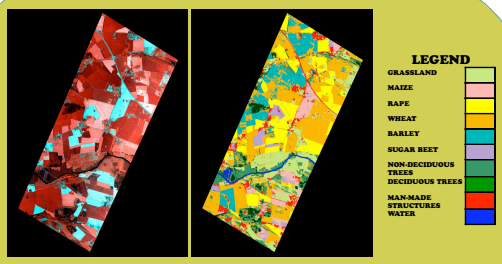
Feature reduction techniques - as compared to feature extraction, feature selection is a more simple and direct approach, and the resulting reduced set of features is easier to interpret. On the other hand, extraction methods are more effective (PCA, NLPCA).

Auto-Associative Neural Networks (AANN)

- Feedforward neural networks
- Non-linear nodes
- Input and target coincide
- Dimensional bottleneck forces the NN to represent significant features in data.
- If the network find an acceptable solution, a good representation of the input must exist in the bottleneck.
- The network can be viewed as two successive functional mapping.



Methodology: a double stage processing has been applied to the data.
•In the first stage a feature reduction has been performed by mean of AANN.
•In the second stage the reduced measurement vector has been used as input to a new neural network scheme for a pixel-based classification procedure.



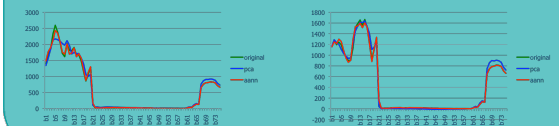
	GRASSLAND	MAIZE	RAPE	WHEAT	BARLEY	SUGAR BEET	NON-DECIDUOUS TREES	DECIDUOUS TREES	MAN-MADE STRUCTURES	WATER	TOTAL
GRASSLAND	**99,47%**	0,00%	**0,15%**	0,00%	0,00%	0,00%	**0,01%**	0,00%	**2,01%**	0,00%	4,27%
MAIZE	0,00%	**99,26%**	0,00%	0,00%	0,00%	**4,88%**	0,00%	0,00%	0,00%	0,00%	5,37%
RAPE	0,00%	0,00%	**99,78%**	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	33,28%
WHEAT	**1,53%**	0,00%	0,07%	**99,36%**	**2,77%**	0,00%	**35,09%**	0,00%	**0,16%**	0,00%	33,15%
BARLEY	0,00%	0,00%	0,00%	**0,67%**	**97,29%**	0,00%	0,00%	0,00%	0,00%	0,00%	10,00%
SUGAR BEET	0,00%	**4,75%**	0,00%	0,00%	0,00%	**95,72%**	0,00%	0,00%	0,00%	0,00%	3,37%
NON-DECIDUOUS TREES	0,00%	0,00%	0,00%	0,07%	**0,26%**	**0,00%**	**64,90%**	**1,59%**	**0,66%**	0,00%	4,04%
DECIDUOUS TREES	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	**0,00%**	**98,41%**	0,00%	**0,04%**	2,37%
MAN-MADE STRUCTURES	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	**96,93%**	0,00%	1,14%
WATER	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	**99,71%**	3,00%
TOTAL	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Overall Accuracy: 98,74%
Kappa Coefficient: 0,9574

Confusion matrix obtained using the ground truth from the ESA-AGRISAR campaign.

A confusion matrix on the same ground truth applied no the result obtained using PCA as input to NN classification show an accuracy of above 67% and also show that some classes are not classified at all.

Comparison: a comparison between the reconstructed spectral signatures using 5 non linear components obtained from AANN and 5 linear components obtained from the PCA. According to the PCA analysis the first 5 components contain almost the 99,9% of the whole statistical information.



Two land cover types are compared (Water, Forest). It can be noticed that the NLPCA is significantly more effective than PCA in encoding the spectral information.

Results: The confusion matrix show an overall accuracy of about 97%. The only confusion elements regards the non-deciduous trees class. In fact, the low accuracy is due to the fact that this class don't came from in situ ground truth, but was detected by a visual inspection of the image, that is subdued to the inspector ability.

Conclusions: in this work a novel approach based on AANNfor the extraction of non-linear principal components from hyper-spectral data has been developed. This approach should be more suitable to eliminate non-linear correlations in the data. A NN algorithm has been exploited to face with the classification task leading to a final single architecture performing the two processing stages: the features extraction aand the classification.
The results show that the features extraction based on auto-associative neural networks outperforms that obtained considering more traditional approach such as PCA. Moreover the reduced vector, used as input to the classification stage , yields land cover maps with high accuracy.