

Remote Sensing Image Classification: A Neuro-fuzzy MCS Approach

B. Uma Shankar, Saroj K. Meher, Ashish Ghosh*, and Lorenzo Bruzzone¹

Machine Intelligence Unit, Indian Statistical Institute
203 B. T. Road, Kolkata 700108, India

¹ Department of Information and Communication Technologies
University of Trento, Via Sommarive, 14, I-38050, Trento, Italy

Abstract. The present article proposes a new neuro-fuzzy-fusion (NFF) method for combining the output of a set of fuzzy classifiers in a multiple classifier system (MCS) framework. In the proposed method the output of a set of classifiers (i.e., fuzzy class labels) are fed as input to a neural network, which performs the fusion task. The proposed fusion technique is tested on a set of remote sensing images and compared with existing techniques. Experimental study revealed the improved classification capability of the NFF based MCS as it yielded consistently better results.

1 Introduction

The objective of designing a pattern classification system is to achieve the best possible performance for the problems at hand. This leads to the development of different classification schemes with different performance levels, and hence they may offer complementary information about the patterns to be classified. This motivated fusing/combining classifiers' outputs for improved performance. The idea is not to rely on a single classifier, rather to use all or some of them for consensus decision making by combining their individual performance. Recently many efforts aimed at it have become popular [1,2,3,4,5,6,7,8,9]. Moreover, the multiple classifier systems (MCSs) are found to be successful with the combination of diverse classifiers. i.e., the classifiers should not commit the same mistake. Further, the performance of an MCS is highly dependent on the combination scheme. Many studies have been published in this area of research, e.g., if only class labels are available a majority voting [10,11] or label ranking [12,13] is used. If continuous outputs like posteriori probabilities are available, an average or some other linear combination can be used [14,15]. If the classifier outputs are interpreted as fuzzy membership values then fuzzy rules [16,17], belief functions and Dempster-Shafer techniques [14,18] can be used for combination.

Classification of land cover regions of remote sensing images is essential for efficient interpretation of them [19,20]. This task is very complex because of low illumination quality and low spatial resolution of remotely placed sensors and

* Corresponding author. ash@isical.ac.in

rapid changes in environmental conditions. Various regions like vegetation, soil, water bodies etc. of a natural scene are often not well separated. Moreover, the gray value assigned to a pixel is an average reflectance of different types of land covers present in the corresponding pixel area. Therefore, a pixel may represent more than one class with varying degree of belonging. Thus assigning unique class label to a pixel with certainty is one of the major problems. Conventional methods cannot deal with this imprecise representation of geological information. Fuzzy set theory introduced in [21] provides a useful technique to allow a pixel to be a member of more than one category or class with graded membership [22]. Many attempts have been made for remote sensing image analysis and classification using fuzzy logic [19,23,24,25,26,27].

We have considered a set of fuzzy classifiers in the design of MCSs in this article. Various existing fuzzy and non-fuzzy fusion methods for combination of classifiers' output are considered and found that the performances are varying with the input data sets and fusion methods. We propose a neuro-fuzzy (NF) fusion method to overcome the risk of selecting a fusion method as neural networks can do this in an adaptive way. The performance is demonstrated on a set of remote sensing images. Experimental study revealed that the MCS with the proposed NF fusion method provided consistently better classification.

2 Fuzzy Classifiers and Combination Methods

A brief description of the four fuzzy classifiers and six existing combination schemes used for the present study is made in the following sections.

Fuzzy k -nearest neighbor (Fk-NN): k -NN is based on the determination of k number of nearest neighbors of a test pattern and assigning it the class label that majority of the neighbors have. Keller *et al.* [28] incorporated the concepts of fuzzy set theory [21] into the k -NN voting procedure and proposed a fuzzy version of k -NN rule. The membership degree of a test pattern \mathbf{x} to class c is calculated as

$$\mu_i(\mathbf{x}) = \frac{\sum_{j=1}^k \mu_{ij} \left(\frac{1}{\|\mathbf{x} - \mathbf{x}_j\|^{2/(m_f-1)}} \right)}{\sum_{j=1}^k \left(\frac{1}{\|\mathbf{x} - \mathbf{x}_j\|^{2/(m_f-1)}} \right)} \quad (1)$$

where $i = 1, 2, \dots, C$ (number of classes), and $j = 1, 2, \dots, k$ (number of nearest neighbors). μ_{ij} is the membership degree of the pattern \mathbf{x}_j from the training set to class i , among the k nearest neighbors of \mathbf{x} . For this study, the algorithm was implemented with $m_f = 2$ and $k = 5$ (selected on the basis of performance).

Fuzzy maximum likelihood (FML): The FML [23] is a fuzzy evaluation of the conventional ML parameters. The mean and variance-covariance matrix

estimated using the fuzzy membership values for each pattern are called *fuzzy mean* and *fuzzy variance-covariance matrix*. The membership function (MF) for class c of a pattern \mathbf{x} can be expressed as

$$f_c(\mathbf{x}) = p_c(\mathbf{x}) / \sum_{j=1}^C p_j(\mathbf{x}) \tag{2}$$

where $p_j(\mathbf{x})$ can be computed as

$$p_j(\mathbf{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_j|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mu_j)^T \Sigma_j^{-1} (\mathbf{x} - \mu_j) \right] \tag{3}$$

with $j = 1, 2, \dots, C$ and D is the dimension of the feature space, assuming a Gaussian distribution for the input data. The fuzzy mean can then be defined as

$$\mu_c = \left(\sum_{i=1}^N f_c(\mathbf{x}_i) \mathbf{x}_i \right) / \left(\sum_{i=1}^N f_c(\mathbf{x}_i) \right) \tag{4}$$

where N is the total number of patterns, f_c is the MF of class c , and \mathbf{x}_i is the i^{th} pattern. The fuzzy variance-covariance matrix can be defined as

$$\Sigma_c = \frac{\sum_{i=1}^N f_c(\mathbf{x}_i) (\mathbf{x}_i - \mu_c) (\mathbf{x}_i - \mu_c)^T}{\sum_{i=1}^N f_c(\mathbf{x}_i)}. \tag{5}$$

For the estimation of the MF in FML, the parameters mean and variance-covariance matrix require the fuzzy representation of the patterns to different classes. Chen [27] described a suitable method which estimates the fuzzy representation of the land covers in an *iterative* manner and does not require the prior information. In this work we have adopted this procedure.

Fuzzy product aggregation reasoning rule (FPARR): The FPARR classification process is performed in three steps [29]. In the first step, it fuzzifies the input feature vector using a π -type MF [30] to get the feature-wise degree of support of a pattern ($\mathbf{x} = [x_1, x_2, \dots, x_d, \dots, x_D]^T$) to all classes. The membership values ($f_{d,c}(x_d)$) thus generated expresses the degree of support of d^{th} feature to c^{th} class. Thus for a pattern \mathbf{x} the membership matrix after the fuzzification process can be expressed as

$$F(\mathbf{x}) = \begin{bmatrix} f_{1,1}(x_1) & f_{1,2}(x_1) & \dots & f_{1,C}(x_1) \\ f_{2,1}(x_2) & f_{2,2}(x_2) & \dots & f_{2,C}(x_2) \\ \dots & \dots & \dots & \dots \\ f_{D,1}(x_D) & f_{D,2}(x_D) & \dots & f_{D,C}(x_D) \end{bmatrix} \tag{6}$$

In the second step, the fuzzified feature values are aggregated using *product reasoning rule* (RR). The RR is operated on the membership matrix column-wise to get the combined membership grade of features to various classes. The

resultant vector thus generated represents the fuzzy classification showing the class belonging.

Fuzzy explicit (FE): The FE classification method [26] also uses three steps. In the first step, it finds the membership matrix for each of the pixels/patterns with a Gaussian MF [26]. Thus a fuzzy membership matrix can be evaluated as in the case of FPARR. The membership matrix is then processed using a *MIN* reasoning rule (RR) in the second step. A rescaling operation is also performed on the output membership values (after applying RR) for all classes.

2.1 Classifier Combination Methods

The fuzzy classifiers' output are combined using a suitable fusion method. For the present study we have considered some of the popular existing techniques. At first the output of the fuzzy classifiers that provide the class belongingness of an input pattern to different classes are arranged in a matrix form defined as decision profile (DP) matrix [31]. It is to be noted that the element of DP matrix need not be fuzzy output only. It could be the labels obtained from the methods providing posterior probability or certainty or possibility values [8]. Mathematically the DP matrix for L classifiers and C classes is defines as

$$DP(\mathbf{x}) = \begin{bmatrix} d_{1,1}(\mathbf{x}) & \dots & d_{1,c}(\mathbf{x}) & \dots & d_{1,C}(\mathbf{x}) \\ \dots & \dots & \dots & \dots & \dots \\ d_{l,1}(\mathbf{x}) & \dots & d_{l,c}(\mathbf{x}) & \dots & d_{l,C}(\mathbf{x}) \\ \dots & \dots & \dots & \dots & \dots \\ d_{L,1}(\mathbf{x}) & \dots & d_{L,c}(\mathbf{x}) & \dots & d_{L,C}(\mathbf{x}) \end{bmatrix}, \quad (7)$$

where $d_{l,c}$ represent the degree of belonging of the pattern assigned by the l^{th} classifier to the c^{th} class.

In the classifier fusion systems, some methods calculate the support for different classes using the corresponding column of DP matrix, regardless of the support for the other classes. This type of fusion methods that use the DP matrix in a class-by-class manner is called class-conscious (CC) combiners. The alternative group is known as class-indifferent (CI). The former uses the context of the DP matrix, i.e., recognizing that a column corresponds to a class, but disregard part of the information with respect to rest of the classes; whereas the CI methods use the whole DP matrix but disregard its context. In the CC group various fuzzy aggregation reasoning rules like *maximum*, *minimum*, *product*, *sum*, *mean* etc. can be applied on each column of the DP matrix. This operation provides a combined output obtained from the aggregation of the classifiers' output for a particular class. Two popular fuzzy combination methods named as *decision template (DT)* and *Dempster-Shafer (DS)* normally work on the DP matrix coming under the CI group. A brief description of the above mentioned fusion methods are given below.

Existing fusion methods: Among the various fusion methods majority voting is the simplest one. In fuzzy aggregation based fusion methods, each column of

the DP matrix is separately combined using various fuzzy aggregation rules like *maximum*, *minimum*, *product*, *sum* and *mean* to get a total support for one class. Maximum of these values for C classes represent the class label for the input pattern [8]. Mathematically the support \bar{D}_j for the class j is defined as

$$\bar{D}_j(\mathbf{x}) = G((d_{1,j}(\mathbf{x}), \dots, d_{L,j}(\mathbf{x})), \dots, d_{L,j}(\mathbf{x})),$$

where \mathbf{x} is the input pattern, $d_{l,j}(\mathbf{x})$ is the membership value of \mathbf{x} obtained by l^{th} classifier for j^{th} class and G is the fuzzy aggregation rule.

The probabilistic product aggregation rule performs fusion using continuous-valued outputs [8]. Let $d_{l,c}(\mathbf{x})$ be the degree of “support” given by classifier l for the class c . Let the feature space be partitioned into L non overlapping subsets which are conditionally independent. Let $P(c)$ denote the prior probability for class c . Then the set of discriminant values proportional to the true posterior probabilities is given by

$$\mu_c(\mathbf{x}) = \frac{\prod_{l=1}^L d_{l,c}(\mathbf{x})}{P(c)^{L-1}}, \quad c = 1, 2, \dots, C. \quad (8)$$

The class label assigned to \mathbf{x} is the highest of $\mu_c(\mathbf{x})$. In this fusion method the fuzzy membership values are assumed to be the probability of a pattern for different classes. Similarly, the fuzzy integral method is performed on each column of the DP matrix to obtain the degree of belonging of each pattern for all classes. In this fusion method, for an input \mathbf{x} , C vectors (of length L) of fuzzy densities are calculated. These values are sorted according to the c^{th} column of the DP matrix. The sorted values are changed iteratively and the final degree of support for class c for an input pattern is calculated.

The next group of fusion methods is known as CI, e.g., DT and DS. The idea of the DT model is to “remember” the most typical DP matrix for each class and then compare it with the current DP(\mathbf{x}) matrix. The closest match will label \mathbf{x} . The $DT_c(\mathbf{x})$ for class c is the average of the DP matrices of the elements of the training set Z in class c . Any kind of similarity measures can be applied for this purpose. A detail description of the method can be found in [31]. In *DS* based fusion method the classifiers’ output are possibility/membership values. Instead of calculating the similarity between the DT_c and DP(\mathbf{x}) matrix, the *DS* algorithm computes the proximity between the DT_c for a class and the output of a classifier, and from this proximity values belief degrees are computed. Based on the belief degrees membership degree for each class is computed for a pattern. The details are available in [14,18].

3 Neuro-fuzzy Combiner

It is well understood that the fuzzy classifiers are suitable for the classification of different ill-defined classes with overlapping boundaries [22]. For remote sensing

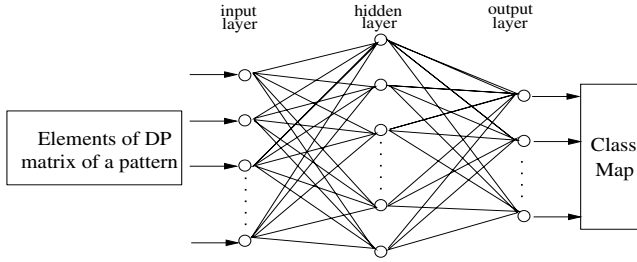


Fig. 1. A three-layer feed-forward neural network

images, classes are normally ill-defined and overlapping. Thus we have used only fuzzy classifiers in the present investigation. We propose a neuro-fuzzy fusion method based MCS that works with fuzzy classifiers, where the output of the classifiers are combined using a neural network (NN) to learn the classes in an iterative way. The activation value of the output neurons show the degree of class labels of the input pattern.

The scheme proposed in the present article falls under the CI category. The elements of the DP matrix are used as input to a three-layered feed forward multi-layer perceptron (MLP) (Fig. 1) which acts as a combiner. Number of input nodes of the NN is equal to the product of number of classifiers used in the MCS and classes present in the data set. Number of output nodes of the NN is equal to the number of classes present in the data set.

Each processing node of MLP, except the input-layer nodes, calculates a weighted sum of the outputs from the nodes in the preceding layer to which it is connected. This weighted sum then passes through a transfer function to derive its own output which is then fed to the nodes in the next layer. Thus, the input and output to node v are obtained as $net_v = \sum_u W_{uv} O_u + bias_v$ and $O_v = S(net_v)$, where w_{uv} is the weight for the connection linking node u to node v , $bias_v$ is the bias value for node v , O_u is the output of node u , and S stands for the activation function (AF) (sigmoid function [32,33]). MLP uses back-propagation (BP) learning algorithm [32,33] for weight updating. The BP algorithm reduce the sum of square error called as cost function (CF), between the actual and desired output of output-layer neurons in a gradient descent manner. The weights are corrected using the following equation:

$$\Delta W_{vu}(n+1) = \alpha \Delta W_{vu}(n) + \eta \delta_v O_u, \quad (9)$$

where n , α , η and δ are the iteration number, momentum parameter, learning rate and node error, respectively. The details of BP algorithm including derivation of the equations can be obtained from [32,33].

The last step of the proposed NFF based MCS system is a hard classification by performing a *MAX* operation to defuzzify the output of the NN. Here the pattern is classified to a class corresponding to the highest node value obtained at output of the NN.

4 Results and Discussion

The proposed scheme has been evaluated using a set of remote sensing images (due to space scarcity we report here results on only two images). Training samples are selected according to a prior assumption of the land cover regions and are used to estimate the parameters of the classifiers. After learning the classifier, it is used to classify the land covers of the whole image.

4.1 Performance Measurement Parameters

Two performance measures described below have been used in the present study.

β index: β is defined [25] as the ratio of the total variation and within-class variation. For a given image and given number of classes, *the higher the homogeneity within the classes, the higher would be the β value.* Mathematically β can be represented as

$$\beta = \left(\sum_{i=1}^C \sum_{j=1}^{M_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}})^2 \right) / \left(\sum_{i=1}^C \sum_{j=1}^{M_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^2 \right), \quad (10)$$

where $\bar{\mathbf{x}}$ is the mean grey value of all the pixels of an image (pattern vector), M_i is the number of pixels in the i^{th} ($i = 1, 2, \dots, C$) class, x_{ij} is the grey value of the j^{th} pixel ($j = 1, 2, \dots, M_i$) in class i , and $\bar{\mathbf{x}}_i$ is the mean of M_i pixel values of the i^{th} class.

Xie-Beni index: The XB index [34] provides a validity criterion based on a function that identifies overall compactness and separation of partition without any assumption to the number of substructures inherent in the data. It is mathematically expressed as the ratio of compactness (θ) and separation (ξ), i.e.,

$$XB = \frac{1}{Z} \frac{\sum_{c=1}^C \sum_{z=1}^Z \mu_{cz}^2 \|V_c - \mathbf{x}_z\|^2}{\min_{c \neq j} \|V_c - V_j\|^2}, \quad (11)$$

where V_c is the centroid of the c^{th} class and \mathbf{x}_z is z^{th} pattern in the data set. Z is the total number of data points in the data set and μ_{cz} is the membership value of the z^{th} pattern to c^{th} class. The *smaller* the XB value, the better is the classification.

4.2 Description of Images

IRS-1A image: The IRS-1A image is obtained from Indian Remote Sensing Satellite [35]. We have used the image taken from the Linear Imaging Self Scanner with spatial resolution of 36.25m x 36.25m and wavelength range of 0.45-0.86 μ m.

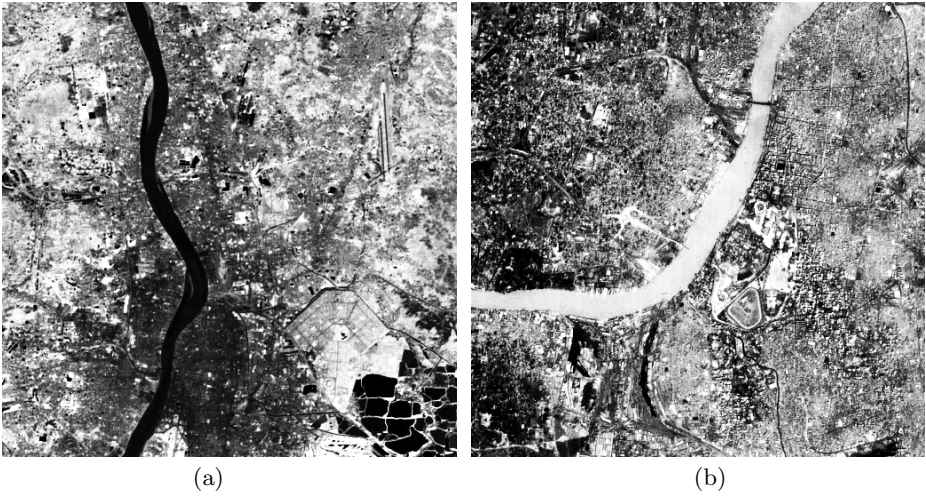


Fig. 2. Original (a) IRS-1A (band-4), and (b) SPOT (band-3) image

The whole spectrum range is decomposed into four spectral bands, namely, blue (band1), green (band2), red (band3) and near infrared (band4) of wavelengths $0.45\text{-}0.52\mu\text{m}$, $0.52\text{-}0.59\mu\text{m}$, $0.62\text{-}0.68\mu\text{m}$, and $0.77\text{-}0.86\mu\text{m}$, respectively. Since the image is poorly illumination, we have presented the enhanced image (band4) in Fig. 2. However, the algorithms are implemented on actual (original) image. The image in Fig. 2a covers an area around the city of Calcutta in the near infrared band having six major land cover classes: *pure water* (PW), *turbid water* (TW), *concrete area* (CA), *habitation* (HAB), *vegetation* (VEG) and *open spaces* (OS). PW class contains pond water, fisheries etc. River water where the soil content is more belong to TW class. CA class consists of buildings, runways, roads, bridges etc. Suburban and rural habitation, where concrete structure are comparatively less come under HAB class. VEG class represents crop and forest areas. OS class contains the barren land.

SPOT image: The SPOT image shown in Fig. 2b is obtained from SPOT satellite (Système Pour d'Observation de la Terre) [20]. The Calcutta image used here has been acquired in the wavelength range of $0.50\text{-}0.89\mu\text{m}$. The whole spectrum range is decomposed into three spectral bands namely, green (band1), red (band2) and near infrared (band3) of wavelengths $0.50\text{-}0.59\mu\text{m}$, $0.61\text{-}0.68\mu\text{m}$, and $0.79\text{-}0.89\mu\text{m}$, respectively. This image has a higher spatial resolution of $20\text{m} \times 20\text{m}$. We have considered the same six classes as in case of IRS-1A image.

4.3 Classification of Remote Sensing Images

Selection of the training samples for all classes are made according to a prior assumption of the land cover regions. These training samples are used to estimate the parameters of the classifiers. After learning the classifier, it is used to classify the land covers of the whole image.

Table 1. β and XB values of individual classification methods

| Sl. No. | Classification method | Image | | | |
|---------|-----------------------|---------------|---------------|---------------|---------------|
| | | IRS-1A | | SPOT | |
| | | β | XB | β | XB |
| 1 | Fk-NN ($k=5$) | 7.0121 | 0.9594 | 6.9212 | 2.5004 |
| 2 | FMLC | 7.0523 | 0.9356 | 6.9896 | 2.4231 |
| 3 | FE | 7.1312 | 0.9112 | 7.0137 | 2.3031 |
| 4 | FPARR | 8.1717 | 0.8310 | 8.1078 | 2.1021 |

Table 2. β and XB values for different combination schemes

| Classifiers for fusion | Fusion method used | | β index | | XB index | |
|--|----------------------------------|---------------|---------------|---------------|---------------|---------------|
| | | | IRS-1A image | SPOT image | IRS-1A image | SPOT image |
| Fk-NN ($k=5$) FMLC FE FPARR | Voting | | 8.3134 | 8.2314 | 0.8211 | 2.1005 |
| | Fuzzy Aggregation reasoning rule | MAX | 0.7903 | 2.1000 | 0.7903 | 2.1000 |
| | | MIN | 8.3213 | 8.5134 | 0.7879 | 1.9733 |
| | | PROD | 8.6217 | 8.6321 | 0.8003 | 2.0178 |
| | | SUM | 8.4312 | 8.3781 | 0.8202 | 2.0013 |
| | | MEAN | 8.2013 | 8.2011 | 0.8201 | 1.9010 |
| | Probabilistic product | | 8.5011 | 8.6005 | 0.7983 | 1.9334 |
| | Fuzzy integral | | 8.5078 | 8.5017 | 0.7710 | 1.9768 |
| | Decision template | | 8.4032 | 8.5712 | 0.7801 | 1.9001 |
| | Dempster-Shafer | | 8.6421 | 8.5312 | 0.7781 | 1.9783 |
| Neuro-fuzzy | | 8.8012 | 8.7763 | 0.7697 | 1.8738 | |

Initially the individual performance of fuzzy classifiers are tested on these images using β and XB indices and depicted in Table 1. It is found that among the four classifiers the FPARR based method is providing the best result.

Further, the validation results (β and XB) produced by MCSs with different combination techniques are provided in Table 2. It is observed that all MCSs with existing fusion methods are providing better results compared to any of the individual classifiers. However, this improvement is not consistent for any of the images. For example, with IRS-1A image, better results are obtained for Dempster-Shafer fusion based MCS with β as a validity measure; whereas fuzzy integral fusion based MCS showed improved performance with XB measure (Table 2). Results are completely different for SPOT image which provided better result with PROD aggregation reasoning rule based MCS in terms of β measure, and DT fusion based MCS in terms of XB measure. Hence, there is a risk in selecting any of these six fusion methods for a particular data set. The performance of the proposed NFF based MCS is then evaluated. The results revealed that the performance is further improved consistently for both the images with respect to the validity measures used here (Table 2). The classified images with this method are shown in Figs. 3a and 3b. It can be seen from these images that all

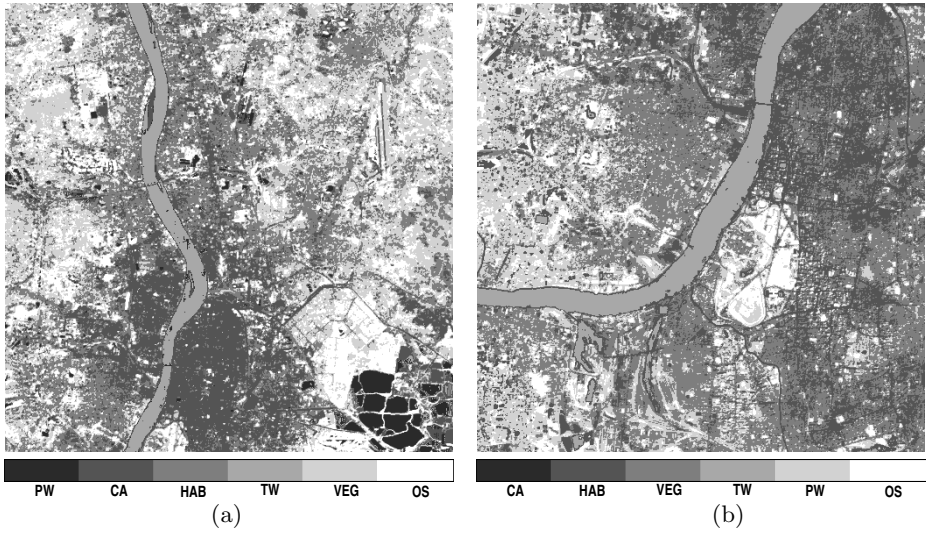


Fig. 3. Classified (a) IRS-1A and (b) SPOT image using proposed fusion based MCS

the classes (PW, TW, CA, HAB, VEG and OS) have come out clearly. Also various structures (like rivers, canals, lakes, roads, bridges, airport runways) present in the images are segregated out properly.

5 Conclusion

A new neuro-fuzzy multiple classifier system (MCS) is presented in this article. Here output of fuzzy classifiers are fed as input to a neural network that acts as a combiner. Performance of the proposed scheme is successfully demonstrated on two remote sensing images, and compared with six existing combination techniques. It is seen that for both the images considered here, the proposed model works well and the improvement is consistent; whereas the results are different for different fusion methods and highly dependent on input data sets.

References

1. Drucker, H., Cortes, C., Jackel, L.D., LeCun, Y., Vapnik, V.: Boosting and other ensemble methods. *Neural Computation* **6** (1994) 1289–1301
2. Bertolami, R., Bunke, H.: Ensemble methods for handwritten text line recognition systems. In: *Proceedings of the 2005 IEEE International Conference on Systems, Man and Cybernetics*. (2005) 2334–2339
3. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20** (1998) 226–239
4. Fumera, G., Roli, F.: A theoretical and experimental analysis of linear combiners for multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27** (2005) 942–956

5. Gutta, S., Huang, J.R.J., Jonathon, P., Wechsler, H.: Mixture of experts for classification of gender, ethnic origin, and pose of human faces. *IEEE Transactions on Neural Networks* **11** (2000) 948–960
6. Moonasar, V., Venayagamoorthy, G.K.: A committee of neural networks for automatic speaker recognition ASR systems. In: *Proceedings of the IJCNN '01. International Joint Conference on Neural Networks. Volume 4.* (2001) 2936–2940
7. Chee, L.P., Harrison, R.F.: Online pattern classification with multiple neural network systems: an experimental study. *IEEE Transactions on Neural Networks* **23** (2003) 235–247
8. Kuncheva, L.I.: *Combining Pattern Classifiers: Methods and Algorithms.* Wiley-Interscience (2004)
9. *Multiple Classifier Systems, International Workshop on MCS. Volume 1857, 2096, 2364, 2709, 3077 and 3541 of LNCS., Springer (2000-2005)*
10. Kimura, F., Shridhar, M.: Handwritten numerical recognition based on multiple algorithms. *Pattern Recognition* **24** (1991) 969–983
11. Franke, J., Mandler, E.: A comparison of two approaches for combining the votes of cooperating classifiers. In: *Proceedings of 11th IAPR International Conference on Pattern Recognition systems.* (1992) 611–614
12. Ho, T.K., Hull, J.J., Srihari, S.N.: Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16** (1994) 66–75
13. Bagui, S.C., Pal, N.R.: A multistage generalization of the rank nearest neighbor classification rule. *Pattern Recognition Letters* **16** (1995) 601–614
14. Xu, L., Krzyzak, A., Suen, C.Y.: Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics* **22** (1992) 418–435
15. Hashem, Schmeiser, B.: Improving model accuracy using optimal linear combinations of trained neural networks. *IEEE Transactions on Neural Networks* **6** (1995) 792–794
16. Cho, S.B., Kim, J.H.: Combining multiple neural networks by fuzzy integral and robust classification. *IEEE Transactions on Systems, Man, and Cybernetics*, **25** (1995) 380–384
17. Cho, S.B., Kim, J.H.: Multiple network fusion using fuzzy logic. *IEEE Transactions on Neural Networks* **6** (1995) 497–501
18. Rogova, G.: Combining the results of several neural network classifiers. *Neural Networks* **7** (1994) 777–781
19. Tso, B., Mather, P.M.: *Classification Methods for Remotely Sensed Data.* Taylor and Francis, London (2001)
20. Richards, J.A., Jia, X.: *Remote Sensing Digital Image Analysis: An Introduction.* 3rd edn. New York: Springer Verlag (1999)
21. Zadeh, L.A.: Fuzzy sets. *Information Control* **8** (1965) 338–353
22. Kuncheva, L.I.: *Fuzzy Classifier Design.* Springer-Verlag (2000)
23. Wang, F.: Fuzzy supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* **28** (1990) 194–201
24. Maselli, F., Rodolfi, A., Copnese, C.: Fuzzy classification of spatially degraded thematic mapper data for the estimation of sub-pixel components. *International Journal of Remote Sensing* **17** (1996) 537–551
25. Pal, S.K., Ghosh, A., Shankar, B.U.: Segmentation of remotely sensed images with fuzzy thresholding, and qualitative evaluation. *International Journal of Remote Sensing* **21** (2000) 2269–2300

26. Melgani, F., Al Hashemy, B.A.R., Taha, S.M.R.: An explicit fuzzy supervised classification method for multispectral remote sensing images. *IEEE Transaction on Geoscience and Remote Sensing* **38** (2000) 287–295
27. Chen, C.F.: Fuzzy training data for fuzzy supervised classification of remotely sensed images. In: *Asian Conference on Remote Sensing (ACRS 1999)*. (1999)
28. Keller, J.M., Gray, M., Givens, J.: A fuzzy k-nearest neighbor algorithm. *IEEE Transactions on Systems, Man and Cybernetics* **15** (1985) 580–585
29. Ghosh, A., Meher, S.K., Shankar, B.U.: Fuzzy supervised classification using aggregation of features. Technical report, MIU/TR-02/2005, Indian Statistical Institute (2005)
30. Pal, S.K., Majumder, D.D.: Fuzzy sets and decision making approaches in vowel and speaker recognition. *IEEE Transactions on Systems, Man and Cybernetics* **7** (1977) 625–629
31. Kuncheva, L.I., Bezdek, J.C., Duin, R.P.W.: Decision templates for multiple classifier fusion: An experimental comparison. *Pattern Recognition* **34** (2001) 299–314
32. Haykin, S.: *Neural Networks: A Comprehensive Foundation*. 2nd edn. Prentice Hall (1998)
33. Lippmann, R.P.: An introduction to computing with neural nets. *IEEE ASSP Magazine* (1987) 4–22
34. Xie, X.L., Beni, G.: A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13** (1991) 841–847
35. NRSA: IRS data users hand book. Technical report (1989) Document No. IRS/NRSA/NDC/HB-02/89.