
Neuro-fuzzy-combiner: an effective multiple classifier system

Ashish Ghosh* and B. Uma Shankar

Machine Intelligence Unit,
Indian Statistical Institute,
203 BT Road, Kolkata 700108, India
E-mail: ash@isical.ac.in
E-mail: uma@isical.ac.in
*Corresponding author

Lorenzo Bruzzone

Department of Information and Communication Technologies,
University of Trento,
Via Sommarive, 14, I-38050, Trento, Italy
E-mail: bruzzone@ing.unitn.it

Saroj K. Meher

Machine Intelligence Unit,
Indian Statistical Institute,
203 BT Road, Kolkata 700108, India
E-mail: saroj_meher@yahoo.com

Abstract: A neuro-fuzzy-combiner (NFC) is proposed to design an efficient multiple classifier system (MCS) with an aim to have an effective solution scheme for difficult classification problems. Although, a number of combiners exist in the literature, they do not provide consistently good performance on different datasets. In this scenario:

- 1 we propose an effective multiple classifier system (MCS) based on NFC that fuses the output of a set of fuzzy classifiers
- 2 conduct an extensive experimental analysis to justify the effectiveness of the proposed NFC.

In the proposed technique, we used a neural network to combine the output of a set of fuzzy classifiers using the principles of neuro-fuzzy hybridisation. The neural combiner adaptively learns its parameters depending on the input data, and thus the output is robust. Superiority of the proposed combiner has been demonstrated experimentally on five standard datasets and two remote sensing images. It performed consistently better than the existing combiners over all the considered datasets.

Keywords: multiple classifier systems; MCSs; fusion of classifiers; combination of classifiers; neuro-fuzzy-combiner; NFC; fuzzy classifier; combination techniques; intermediate feature space; knowledge engineering; soft data paradigm.

Reference to this paper should be made as follows: Ghosh, A., Shankar, B.U., Bruzzone, L. and Meher, S.K. (2010) 'Neuro-fuzzy-combiner: an effective multiple classifier system', *Int. J. Knowledge Engineering and Soft Data Paradigms*, Vol. 2, No. 2, pp.107–129.

Biographical notes: Ashish Ghosh is a Professor of the Machine Intelligence Unit at the Indian Statistical Institute, Calcutta. He received the prestigious and most coveted Young Scientists Award in Engineering Sciences from the Indian National Science Academy in 1995 and in Computer Science from the Indian Science Congress Association in 1992. He has been selected as an Associate of the Indian Academy of Sciences, Bangalore in 1997. His research interests include pattern recognition and machine learning, data mining, image analysis, remotely sensed image analysis, soft computing, fuzzy sets and uncertainty analysis, neural networks, evolutionary computation and bioinformatics. He has already published more than 100 research papers in internationally reputed journals and referred conferences, and is acting as a member of the Editorial Board/Associate Editor of various international journals. He is the in-charge and a member of the founding team that established a National Center for Soft Computing Research at the Indian Statistical Institute, Kolkata in 2004 with funding from the Department of Science and Technology (DST), Government of India.

B. Uma Shankar is an Assistant Professor of Machine Intelligence Unit at the Indian Statistical Institute, Kolkata, India. His research interests are in the areas of image processing, pattern recognition, machine learning, neural networks, fuzzy sets, rough sets, soft computing and analysis of remote sensing images. He has authored and co-authored more than 30 scientific articles in international journals and referred conferences. He is a Life-Member of IUPARAI (an affiliated body of IAPR) and a member of IRSC and IEEE.

Lorenzo Bruzzone is a Full Professor of Telecommunications and the Head of the Remote Sensing Laboratory at the University of Trento, Italy. His research interests are in the areas of remote sensing, radar, signal processing and pattern recognition. He is the author and co-author of 90 scientific publications in referred international journals, more than 140 papers in conference proceedings, and 13 book chapters. He is Editor/Co-Editor of ten books/conference proceedings. He was a recipient of many awards and scientific recognitions and was involved in the ideation and organisation of many international conferences and projects. He is also a Fellow of IEEE.

Saroj K. Meher is presently working as a Visiting Faculty in Center for Soft Computing Research, A National Facility, Indian Statistical Institute, India. He received his PhD from National Institute of Technology, Rourkela, India, in 2003. His research interest includes granular computing, pattern recognition, soft computing methods and digital signal processing. He has published many research articles in internationally reputed journals and refereed conferences.

1 Introduction

The process of consulting several experts before making a final decision is perhaps the second nature of human being; yet, the extensive benefits of such a metaphor in automated decision-making process have only been discovered recently by pattern recognition community (Polikar, 2006). This metaphor is being used to design *multiple*

classifier systems (MCSs). The task of a pattern classifier is to assign a pattern to its actual class (Theodoridis and Koutroumbas, 2006). Even if different classifiers are trained under the same environment, often they assign contradictory labels to unknown patterns. This motivated the design of a classifier system through fusion/combination of multiple classifiers. The idea is not to rely on a single classifier; instead, all the classifiers or their subsets are used for decision-making by combining their individual performances. This is because group decisions are generally better than any individual's decision. The objective of MCS is to exploit the complementary discriminatory ability of different classifiers for achieving improved learning (training) and better generalisation (testing) (Kittler, 1998).

Research in the area of machine learning has achieved significant height with the concept of learning from labelled instances. Although many efficient algorithms have been proposed for this, their utilities have been restricted to simple problems only. Furthermore, numerous results suggest that learning more difficult concepts tend to be extremely difficult with this idea. One of the research directions that have evolved to address these issues is the construction of MCSs. A good understanding of how to build more sophisticated MCSs and exploit various possibilities of extracting information from the environment may help us to move closer to achieve the original intent of machine learning [i.e., to automate the knowledge acquisition process (*International Workshop on Multiple Classifier Systems*, 2000–2005 and 2007)]. Note that MCS has two major goals:

- 1 increasing the classification accuracy
- 2 increasing the robustness of the system in terms of generalisation ability on unlabeled patterns.

Combination of multiple classifiers may be treated as one of the most exciting advancements in pattern recognition over the last decade. MCSs, also known by other names like ensemble learning (Hansen and Salamon, 1990), classifier combination (Lee and Srihari, 1995a), mixture of experts (Guttag et al., 2000) etc., have been demonstrated to offer means for enhancing the performance of pattern recognition systems (*International Workshop on Multiple Classifier Systems*, 2000–2005 and 2007; Bertolami and Bunke, 2005; Cho and Kim, 1995a, 1995b; Drucker et al., 1994; Fumera and Roli, 2005; Guttag et al., 2000; Hashem and Schmeiser, 1995; Ho et al., 1994; Kimura and Shridhar, 1991; Kittler et al., 1998; Kuncheva, 2004; Kuncheva et al., 2001; Rogova, 1994; Xu et al., 1992). Improvement in performance due to classifier combination is reviewed and presented by Kittler (1997) and Jain et al. (2000). Witten and Frank (2005) described some of the methods for combining multiple models with applications to machine learning and data mining. The book by Kuncheva (2004) is an attempt to provide a comprehensive survey of this fast growing field. It covers a broad range of methodologies, algorithms, and theories with case studies and real life applications. A more recent survey by Polikar (2006) describes the design, implementation and applications of many of the existing ensemble systems in decision-making process.

Although many efforts aimed at MCSs have become popular, the performance of an MCS is highly dependent on the combination technique (Kuncheva, 2004) and the input dataset. The use of multiple classifiers led to increase in computational complexity with significant increase in cost of computation and response time. To overcome these constraints a trade-off between the classification accuracy and the cost of computation needs to be carefully considered in real-world applications. Demir and Alpaydin (2005)

evaluated the cost of combining classifiers and proposed *cost-conscious classifier ensembles*.

Normally, a set of classifiers are used in the design of an MCS for combining the output through some linear or non-linear method. The process of combination can be anything starting from simple voting or linear combination to adaptive learning. The weighted averaging based combiners mainly use linear regression procedure to fix the weights (Hashem, 1997; Jacobs, 1995; Ting and Witten, 1999). In a recent work by Raudys (2006a, 2006b) on trainable fusion rules, focus is given on the effect of finite samples (small and large). Although they demonstrated that non-trainable fusion rules can outperform more sophisticated trainable ones for small samples, they opposed fixed fusion rules over trainable ones and demonstrated situations where weighted averaging based fusion methods can outperform simple averaging based fusion methods for large samples.

The improvement and modification of MCSs technique can be carried out at any of the following four levels (Kuncheva, 2004):

- a combination level
- b classifier level
- c feature level
- d data level.

We mainly consider combination level issues in the present study. Some experimental studies have already been carried out in this regard (*International Workshop on Multiple Classifier Systems*, 2000–2005 and 2007; Kittler, 2002; Kuncheva, 2004) and it is found that there are difficulties in selecting a suitable combiner for a particular problem. For example, if only class labels are available a majority voting (Kimura and Shridhar, 1991) or label ranking (Ho et al., 1994) may be used. If continuous outputs like posterior probabilities are available, an average or some other linear combination techniques can be adopted (Hashem and Schmeiser, 1995; Xu et al., 1992). If the classifier output is interpreted as fuzzy membership values (degree of support, certainty, or possibility), belief values or evidence, then fuzzy rules (Cho and Kim, 1995b), belief functions and Dempster-Shafer (DS) method (Rogova, 1994; Xu et al., 1992) can be exploited for combination. Similarly, for some other applications probabilistic product (PP) (Kuncheva, 2004) or decision template (DT) (Kuncheva et al., 2001) can be applied on the output of the individual classifiers.

Another way to design a combiner is to train it as a classifier using the output of base classifiers (Kittler, 1998, 2002; Krogh and Vedelsby, 1995; Kuncheva, 2004; Wolpert, 1992). In this regard, a neural network (NN) has also been used as a learnable combiner (El-Melegy and Ahmed, 2007; Gader et al., 1996; Huang and Suen, 1994; Huang et al., 1995; Lee, 1995; Lee and Srihari, 1995a, 1995b; Liu et al., 1994; Mahmoud et al., 2004). Here, it is necessary to mention the work of Kittler (1998, 2002) with respect to trainable classifier combiner. According to the author when the linear or non-linear combination functions are estimated through a training process the distinctions between the two scenarios, fusion of opinions based on identical and on non-identical representations, fade away. Kittler (2002) also stated “this probably explains the success of many heuristics combination strategies that have been suggested in the literature without any concerns about the underline theory”.

The use of NN for adaptive learning in the framework of a combiner was studied in the past (El-Melegy and Ahmed, 2007; Gader et al., 1996; Huang and Suen, 1994; Huang et al., 1995; Kuncheva, 2003; Lee, 1995; Lee and Srihari, 1995a, 1995b; Liu et al., 1994; Mahmoud et al., 2004; Meher et al., 2006; Shankar et al., 2006). Lee and Srihari (1995a) used a NN called decision combination neural network (DCNN) for combining classifiers utilising confidence information for handling mixed type of classifiers. Further modifications of this method are available in Lee (1995), and Lee and Srihari (1995b). Huang et al. (1995) and Liu et al. (1994) presented a combination of multiple expert (CME) method using NNs, which can efficiently and effectively combine several classifiers. Huang and Suen (1994) tried to transform measurement values of different meaning and scales (features) into those of the same meaning and scale, and used a NN to combine the transformed data. In Gader et al. (1996), the authors made a comparison of different combiners (including a neural one) for the problem of word recognition. In El-Melegy and Ahmed (2007), and Mahmoud et al. (2004), the authors gave a comparison of the performance of several statistical and neural fusion methods. They concluded that even if all individual classifiers are optimised, decision fusion approaches can improve the overall performance. However, the benefit from decision fusion may be limited when there is a small number of training data, or when the classification accuracy of an individual classifier is sufficiently high. In Kuncheva (2003), gave a review of different combination techniques of classifiers in soft-computing paradigm. A combiner to be termed as soft or not will depend on the problem it is used for and the solutions it produces. Therefore, one can use any classifier as a combiner for combining the output of base classifiers. Our objective here is to combine the soft output of base classifiers and the combiner also should provide soft output. There are many learning based combiners studied in literature (Kuncheva, 2004). In the present study, we have compared our proposed neuro-fuzzy-combiner (NFC) with some relevant combiners, which include voting, fuzzy aggregation, PP, fuzzy integral, DT and DS theory for aggregation.

Though there exists a strong aspiration for the theoretical studies on multiple classifier combination techniques, even now they rely mainly on heuristic and empirical solutions (Kuncheva, 2004). It is expected that the MCS can provide better results than a single classifier (*International Workshop on Multiple Classifier Systems*, 2000–2005 and 2007; Kuncheva, 2004; Polikar, 2006). MCSs are found to be successful with the combination of diverse base classifiers, i.e., the classifiers which do not commit the same mistake. Many studies have been published in this area of research to compare various combiners (Alkoot and Kittler, 1999; Battiti and Colla, 1994; Fumera and Roli, 2005; Jacobs, 1995; Kittler et al., 1998; Kuncheva, 2002a, 2002b; Tax et al., 2000). Our study with six widely used combiners also corroborates this finding. In general, there is no single algorithm which can solve all the problems optimally. Similarly, *for combination methods also there is no best combiner for all problems*. So there is a need to find a combiner which can perform consistently well and may not depend on the dataset and selection of base classifiers.

The aim of the present work is to find a combiner which is better than the existing popular combiners and is not much dependent on the input datasets. In this regard, we have used a neuro-fuzzy framework which has already been applied successfully to several other real-life problems (Bunke and Kandel, 2000; Jang et al., 1996; Lin and Lee, 1996; Mitra and Hayashi, 2000; Pal and Mitra, 1999). The proposed methodology combines the output of fuzzy classifiers using a ‘learnable’ classifier (a NN).

Performance of the proposed neuro-fuzzy MCS is found to be consistently better than several other fuzzy combination rules for the considered datasets. Here, we propose an adaptive combiner which learns its parameters by using the output of a set of fuzzy classifiers as input. In other words, the estimated parameters extract general properties of the input data. The proposed method uses the principle of neuro-fuzzy hybridisation to combine the output of fuzzy classifiers. The combination also produces fuzzy output. It may be noted that the proposed NFC is conceptually different from the existing neural-network based combiners, because it uses the NF framework and takes output from fuzzy classifiers only as input and provide fuzzy output. As to the knowledge of the authors, this has not been explored in the past. This fuzzy output has an additional advantage of using it at a different level/stage of processing (e.g., in remote sensing image analysis these fuzzy outputs are more useful for detection of linear objects, like roads structures). Here, we demonstrate its performance on a set of standard datasets and a set of remote sensing images. The proposed method is compared with six popular methods (Kuncheva, 2004). Most of which combine the output of fuzzy classifiers by using other tools (not NN) and provides fuzzy output. Experimental study revealed that the MCS with the proposed NFC provided consistently more accurate and robust classification results.

The rest of the article is organised as follows: Section 2 provides a brief description of four fuzzy classifiers and six existing combination techniques used in the present study. The proposed NFC is described in Section 3. Section 4 presents detailed discussion of the results obtained on various datasets. Concluding remarks are given in Section 5.

2 Fuzzy classifiers and combination methods

The problem of pattern classification is a complex task, particularly when it is modelled according to a multi-stage process like in MCS. For this kind of methods fuzzy output at the first stage are very useful. A soft decision at the intermediate stages about the considered samples provides more flexibility in the final decision process. In addition, fuzzy set theory (Zadeh, 1965) provides a useful methodological tool that allow pattern to be a member of more than one category (Kuncheva, 2000). In this way, we are not committing ourselves to a particular decision rather the degree of commitment is measured in terms of membership to different classes. Each of the base (fuzzy) classifiers provides graded membership value for a pattern to belong to different classes, which can be combined in an MCS to get a final decision. As mentioned earlier, fuzzy output are normally combined by fuzzy rules, belief functions and DS techniques.

2.1 Fuzzy classifiers

We have considered four fuzzy classifiers for the present investigation. These are:

- 1 fuzzy k -nearest neighbour (Fk -NN) (Keller et al., 1985)
- 2 fuzzy maximum likelihood (FML) (Wang, 1990)
- 3 fuzzy product aggregation reasoning rule (FPARR) (Ghosh et al., 2008)
- 4 fuzzy explicit (FE) (Melgani et al., 2000).

Brief descriptions of these four fuzzy classifiers are provided in the following sections.

2.1.1 Fuzzy k -nearest neighbour

The k -NN (k -nearest neighbour) classifier is based on the determination of k number of nearest neighbours of a test pattern and assigning it the class label that majority of the neighbours have (Theodoridis and Koutroumbas, 2006). Keller et al. (1985) incorporated the concepts of fuzzy set theory (Zadeh, 1965) into the k -NN voting procedure and proposed a fuzzy version of k -NN rule. The membership degree of a test pattern x to class c is calculated as:

$$\mu_c(x) = \frac{\sum_{j=1}^k \mu_{cj} \left(\frac{1}{\|x - x_j\|^{2/(m_f-1)}} \right)}{\sum_{j=1}^k \left(\frac{1}{\|x - x_j\|^{2/(m_f-1)}} \right)} \quad (1)$$

where $c = 1, 2, \dots, C$ (number of classes), and $j = 1, 2, \dots, k$ (number of nearest neighbours), and m_f is the fuzzifier. μ_{cj} is the membership degree of the training pattern x_j (from among the k nearest neighbours of x) to class c .

2.1.2 Fuzzy maximum likelihood

The FML (Wang, 1990) is a fuzzy evaluation of the conventional maximum likelihood parameters. The mean vectors and covariance matrices estimated using the fuzzy membership values for each pattern are called *fuzzy mean vectors* and *fuzzy covariance matrices*. The membership value of a pattern x for the c th class can be expressed as:

$$\mu_c(x) = p_c(x) / \sum_{j=1}^C p_j(x) \quad (2)$$

with $j = 1, 2, \dots, c, \dots, C$. Assuming Gaussian distribution for the input data, the probability density function $p_j(x)$ for the j th class can be computed as:

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

where D is the dimension of the feature space, $|\Sigma_j|$ is the determinant of Σ_j (the covariance matrix for j th class), and \bar{x}_j is the mean of patterns in class j . The fuzzy mean vector of the c th class can then be defined as:

$$\bar{x}_c = \left(\sum_{i=1}^{M_c} \mu_c(x_i) x_i \right) / \left(\sum_{i=1}^{M_c} \mu_c(x_i) \right) \quad (4)$$

where M_c is the total number of patterns in the c th class, μ_c is the membership value to class c , and x_i is the i th pattern. The fuzzy covariance matrix for c th class is Σ_c , which can be defined as:

$$\Sigma_c = \frac{\sum_{i=1}^{M_c} \mu_c(x_i)(x_i - \bar{x}_c)(x_i - \bar{x}_c)^T}{\sum_{i=1}^{M_c} \mu_c(x_i)}. \quad (5)$$

For the estimation of the membership-function in FML, the parameters mean vectors and covariance matrices require the fuzzy representation of the patterns to different classes. Chen (1999/2000) described a method that estimates the fuzzy representation of the patterns in an *iterative* manner and does not require any prior information.

2.1.3 Fuzzy product aggregation reasoning rule

The FPARR classification process is performed in three steps (Ghosh et al., 2008). In the first step, it fuzzifies the input feature vector using a π -type membership-function (Pal and Majumder, 1977) to get the feature-wise degree of support of a pattern ($x = [x_1, x_2, \dots, x_d, \dots, x_D]^T$) to all classes. The membership values ($\mu_{d,c}(x_d)$) thus generated expresses the degree of support of d th feature to c th class. Thus, for a pattern x the membership matrix $F(x)$, after the fuzzification process, can be written as:

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

In the second step, the fuzzified feature values are aggregated using the *product* (PROD) 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 support for various classes (class belonging). The third step is used for defuzzification. Note that this step will not be required with the considered MCSs, except in case of majority voting.

2.1.4 Fuzzy explicit

The FE classification method (Melgani et al., 2000) also uses only three steps similar to FPARR. In the first step, it finds the membership matrix for each of the patterns with a Gaussian membership-function (Melgani et al., 2000). Thus, a fuzzy membership matrix can be evaluated as in the case of FPARR. The membership matrix is then processed using a *MIN* (minimum) RR in the second step. As in FPARR, the third step is used for defuzzification and will not be required with the considered MCSs, except in case of majority voting.

Note that, in general a rescaling operation is also performed on the output of the classifiers to make it uniform before applying any combination rule.

2.2 Classifier combination methods

At first the output of the fuzzy classifiers (which provide membership of an input pattern to different classes) are arranged in a matrix form defined as decision profile (DP) matrix

(Kuncheva et al., 2001). It is to be noted that the elements of DP matrix do not need to be fuzzy output only. They could be the labels obtained from the methods providing posterior probability or certainty or possibility values (Kuncheva, 2004) also. Mathematically, the DP matrix for L classifiers and C classes is defined as:

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

where $d_{l,c}$ represents the degree of belonging of a pattern assigned by the l th classifier to the c th class. In the present study all the elements of the DP matrix are membership values obtained from different fuzzy classifiers to various classes.

In a classifier combination system, 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 combiners that use the DP matrix in a class-by-class manner are 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., recognising that a column corresponds to a class, but disregards part of the information with respect to rest of the classes; whereas the CI methods use the whole DP matrix and disregards its context. It may be noted here that we can ignore the context of DP and treat the elements of DP as features in a new feature space, which is called the *intermediate feature space* (Kuncheva, 2004). In the CC group various fuzzy aggregation RRs, like *maximum*, *minimum*, *product*, *sum*, *mean*, etc. can be applied to each column of the DP matrix (Kuncheva, 2004). 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 *DT* (Kuncheva, 2001; Kuncheva et al., 2001) and *DS* method (Rogova, 1994; Xu et al., 1992) normally work on the DP matrix coming under the CI group. A brief description of these combiners is given below.

2.2.1 Class conscious combiners

Among the various combiners majority voting is the simplest. In fuzzy aggregation based combiners, 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 each class (Kuncheva, 2004). Maximum over these values for C classes provides the class label for the input pattern (Kuncheva, 2004). Mathematically, the support \bar{D}_c for the class $c (= 1, 2, \dots, C)$ is defined as:

$$\bar{D}_c(x) = G(d_{1,c}(x), \dots, d_{l,c}(x), \dots, d_{L,c}(x)),$$

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

The *PP* aggregation rule combines continuous-valued output (Kuncheva, 2004). Let $d_{l,c}(x)$ be the degree of 'support' given by classifier l for 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:

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

The class label assigned to x is the highest of $\mu_c(x)$.

Similarly, the *fuzzy integral* (FI) (Cho and Kim, 1995a; Kuncheva, 2004; Tahani and Keller, 1990) is performed on each column of the DP matrix to obtain the degree of membership of each pattern to all classes. In this combiner, for an input 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 each input pattern is calculated.

2.2.2 Class indifferent combiners

The next group of combiners (e.g., DT and DS) is known as CI. The idea of the *DT model* is to ‘remember’ the most typical DP matrix for each class and then compare it with the current $DP(x)$ matrix. The closest match will label x . The $DT_c(x)$ for class c is the average of the DP matrices of the elements of the training set in class c . Any kind of similarity measure can be applied for this purpose. Detail description of the method can be found in Kuncheva et al. (2001). This calculation is equivalent to applying the nearest mean classifier in the intermediate feature space (Kuncheva, 2004). In *DS based MCSs*, the base classifiers’ output are possibility/membership values. Instead of calculating the similarity between the DT_c and $DP(x)$ 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. Let the C DTs be DT_1, DT_2, \dots, DT_C computed from the data. Instead of calculating similarity between the decision template DT_c and the decision profile $DP(x)$, the following steps are carried out. Let DT_c^l denote the l th row (classifier) of decision template DT_c . $D_l(x)$ denotes the output (membership value) of D_l ; that is $D_l(x) = [d_{l,1}(x), d_{l,1}(x), \dots, d_{l,c}(x), \dots, d_{l,C}(x)]$, the l th row of the DP matrix $DP(x)$. We calculate the ‘proximity’ Φ between DT_c^l and the output of classifier D_l for the input x as:

$$\Phi_{c,l}(x) = \frac{\left(1 + \|DT_c^l - D_l(x)\|^2\right)^{-1}}{\sum_{j=1}^C \left(1 + \|DT_j^l - D_l(x)\|^2\right)^{-1}}, \quad (9)$$

where $\| \cdot \|$ is any matrix norm (e.g., Euclidean norm between two vectors). Thus, for each DT we have L (number of classifier) proximities. Using equation (9), we can calculate for every class $c = (1, 2, \dots, C)$ and for every classifier $l = (1, 2, \dots, L)$, the *belief degrees* as follows:

$$b_c(D_l(x)) = \frac{\Phi_{c,l}(x) \prod_{j=1, j \neq c}^C (1 - \Phi_{j,l}(x))}{1 - \Phi_{c,l}(x) \left[1 - \prod_{j=1, j \neq c}^C (1 - \Phi_{j,l}(x)) \right]} \quad (10)$$

where $b_c(D_l(x))$ is the belief degree with respect to the c th class and the l th classifier and the final degree of the support for the c th class is given by:

$$\mu_c(x) = \mathcal{K} \prod_{l=1}^L b_c(D_l(x)), \quad c = 1, 2, \dots, C \text{ and } \mathcal{K} \text{ is a normalising constant.} \quad (11)$$

The details on DS combination method can be found in Kuncheva (2004), Rogova (1994) and Xu et al. (1992).

3 Proposed NFC

It is well understood that the fuzzy classifiers are suitable for classification of different ill-defined classes with overlapping boundaries (Kuncheva, 2000) (e.g., in remote sensing images where classes are normally ill-defined and overlapping). In many real-life problems fuzzy classifiers are proved successful because of their ability to assign soft class labels to sample patterns (Kuncheva, 2000). Thus, we have used only fuzzy classifiers in the present investigation. In this context, we propose a NFC based MCS that works with fuzzy classifiers where the output of the classifiers are combined using a NN to learn the classes in an iterative way. The activation value of the output neurons show the degree of class belonging of the input pattern. Here, our aim is to show some advantages of NFC where the fuzzy class label of patterns generated by fuzzy classifiers are fused by a MLP. Note that, it is always better to take a decision at the final stage, rather than taking decisions at an intermediate stage. The fuzzy output of the fuzzy classifiers allow us not to commit any decision at the intermediate stages; and these output are used as intermediate features to a NN where the output of the NN is also a continuous value, providing ample opportunity to use the final output for higher level processing (e.g., image analysis). In the proposed model we tried to exploit this characteristic. Performance of the proposed neuro-fuzzy MCS was found to be consistently better than several other combination rules for the considered datasets.

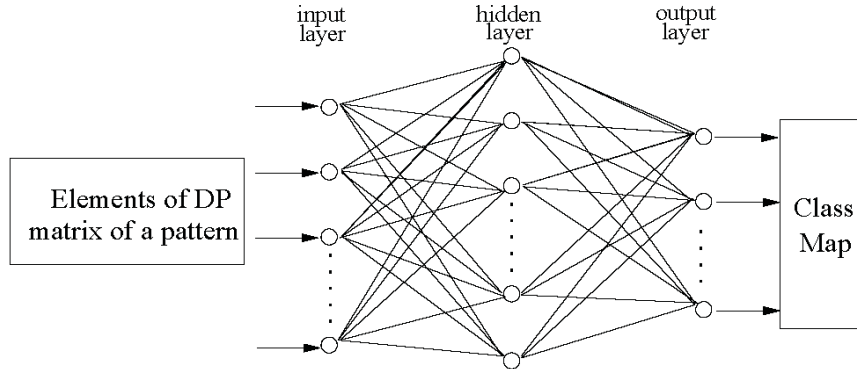
Designing a well performing fuzzy system is not always an easy task. The problem of finding appropriate membership function and fuzzy rules is often a tiring process of trial and error. Therefore, the idea of applying learning algorithms to fuzzy systems was considered in the past. There are various kinds of adaptive models used in the learning approaches, one of them is neural-network leading to a new technique, known as neuro-fuzzy technique. This became a popular choice for various reasons. The learning capabilities of NNs made them a prime target for combination with fuzzy systems in order to automate or support the process of developing a fuzzy system for a given task. A neuro-fuzzy system is a fuzzy system trained by a neural learning algorithm (usually). The learning procedure operates on local information and causes only local modifications in the underlying fuzzy system (Bunke and Kandel, 2000; Jang et al., 1996; Lin and Lee, 1996; Mitra and Hayashi, 2000; Pal and Mitra, 1999). The learning process is data driven.

Using the fuzzy sets encoded in the connection weights, the rules for the fuzzy system are generated through the learning process.

The technique proposed in the present paper falls under the CI category. The elements of the DP matrix are used as input to a feed forward multi-layer perceptron (MLP) (Figure 1) which acts as the 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 dataset. Number of output nodes of the NN is equal to the number of classes present in the dataset. As a rule of thumb, the number of hidden nodes in the NN is the square root of the product of number of input nodes and the number of output nodes.

Each processing node of MLP, except the input-layer nodes, calculates a weighted sum of the output 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 a sigmoid function (Haykin, 1998). MLP uses back-propagation learning algorithm (Haykin, 1998) for weight updating. The back-propagation learning algorithm reduces the sum of square error called cost function, between the actual and desired output of output-layer neurons in a gradient descent manner.

Figure 1 A three-layer feed-forward neural network



The weights are corrected using the following equation:

$$\Delta W_{vu}(n+1) = \alpha \Delta W_{vu}(n) + \eta \delta_v O_u, \quad (12)$$

where n, α, η and δ are the iteration number, momentum parameter, learning rate and node error, respectively (appropriate parameters are chosen using prior knowledge). The details of back-propagation learning algorithm including derivation of the equations can be obtained from Haykin (1998).

The last step of the proposed NFC based MCS system is hard classification by performing a MAX (maximum) operation to defuzzify the output of the NN. Here the input pattern is assigned to a class corresponding to the highest node value obtained at output of the NN. However, without defuzzification also the output can be used for further processing, e.g., in image analysis, the fuzzy classification can be used for mixed

type pixels for making final decision using other information (i.e., mid and high level vision applications).

4 Results and discussion

In the present study, we have considered five standard datasets and two remote sensing images.

4.1 Results on standard datasets

The set of five datasets contains *pima* (Asuncion and Newman, 2007), *vowel* (Pal and Majumder, 1977), *phoneme* (Elena database), *caldonazzo* and *satimage* (Elena database), of which two datasets (*caldonazzo* and *satimage*) are from remote sensing images. A brief description of these five datasets is given in Table 1.

Table 1 Brief description of the datasets

<i>Number of dataset</i>	<i>Number of classes</i>	<i>Number of features available</i>	<i>Number of features used</i>	<i>Number of patterns</i>
Pima	2	8	8	768
Vowel	6	3	3	871
Phoneme	2	5	5	5,404
Caldonazzo	6	7	6*	3,894
Satimage	6	36	4*	6,435

Note: *As suggested by the contributor of the dataset.

Selection of the training samples for all classes has been made by dividing them into two parts. First part is taken for estimation of the parameters of the classifiers and the combiner together (training data). The second part is taken for testing the performance of MCSs (test data).

In the present study, experiments were performed with 10%, 30% and 50% training samples drawn randomly from the whole dataset and the remaining 90%, 70% and 50% are used for testing. The same percentage of data is drawn from each class. Results were found to be very consistent and results for one such simulation with 30% training data is presented for a typical illustration. This also helped us to show that size of the training data does not affect the performance much. The results show that the proposed combiner performs consistently better than the existing combiners with the considered datasets. In other words, it is not dependent on the input data (one of the main drawbacks of the existing combiners).

Initially, the individual accuracies of four classifiers are tested with considered datasets (see Table 2). It is found that on all datasets FPARR based classifier provided the best result compared to the other three algorithms. This is true for all the datasets. A comparative analysis of the proposed NFC (with these fuzzy classifiers as base classifiers) is made with similar existing six combiners. These are majority voting (Kimura and Shridhar, 1991), fuzzy aggregation operators (Cho and Kim, 1995b; Hashem and Schmeiser, 1995; Xu et al., 1992), PP (Kuncheva, 2004), fuzzy integral

(Cho and Kim, 1995a; Kuncheva, 2004; Tahani and Keller, 1990), DT (Kuncheva, 2001; Kuncheva et al., 2001) and DS method (Rogova, 1994; Xu et al., 1992).

Table 2 Overall classification accuracy with different classification methods

Classification method	Overall accuracy (%)				
	<i>Pima</i>	<i>Vowel</i>	<i>Phoneme</i>	<i>Caldonazzo</i>	<i>Satimage</i>
Fk-NN	71.76	75.88	74.01	74.74	77.25
FMLC	72.02	76.05	74.67	75.06	77.89
FE	72.98	76.80	74.65	75.88	78.11
FPARR	73.44	77.29	77.44	76.57	80.46

Classification accuracies obtained by these MCSs are reported in Table 3. It is observed that on all datasets the MCSs are providing better classification results, as expected, compared to any of the individual classifiers. However, from the analysis of results one can observe that for different combiners the accuracies are varying. Moreover, the improvement is also dependent on the input datasets (e.g., with *pima*, *vowel*, *phoneme*, *caldonazzo* and *satimage* datasets, better results are obtained with MCSs based on *DS*, *DT*, *product* aggregation RR, *minimum* aggregation RR and *fuzzy integral*, respectively (Table 3). Hence, it is difficult to choose a combiner for a particular dataset. However, with the proposed NF combiner based MCS, the accuracies are improved and found to be superior for all datasets compared to the results obtained with any of the six previously discussed existing combiner based MCSs.

Table 3 Overall accuracy of MCSs with different combination techniques

Combination technique	Overall accuracy (%)					
	<i>Pima</i>	<i>Vowel</i>	<i>Phoneme</i>	<i>Caldonazzo</i>	<i>Satimage</i>	
Voting	74.52	78.71	77.76	78.66	80.67	
Fuzzy aggregation reasoning rule	Maximum	75.25	78.63	77.60	77.83	80.01
	Minimum	75.51	79.42	78.75	78.91	81.70
	Product	75.40	79.37	79.87	78.47	82.16
	Sum	74.89	78.89	77.83	77.84	80.93
	Mean	74.48	77.69	77.47	77.11	80.46
Probabilistic product	75.45	79.23	79.23	78.71	81.34	
Fuzzy integral	74.71	78.68	79.30	76.42	82.76	
Decision template	74.62	79.76	77.95	78.01	82.34	
Dempster-Shafer	75.69	78.97	79.01	77.92	82.13	
<i>NFC</i>	77.21	80.11	81.75	79.47	83.41	

4.2 Results on remote sensing images

Classification of land cover regions of remote sensing images is essential for efficient interpretation of them (Richards and Jia, 1999; Tso and Mather, 2001). This task is very complex because of low illumination quality and low spatial resolution of satellite sensors and rapid changes in environmental conditions. Fuzzy set theory (Zadeh, 1965) provides a useful technique to allow a pixel to be a member of more than one category or

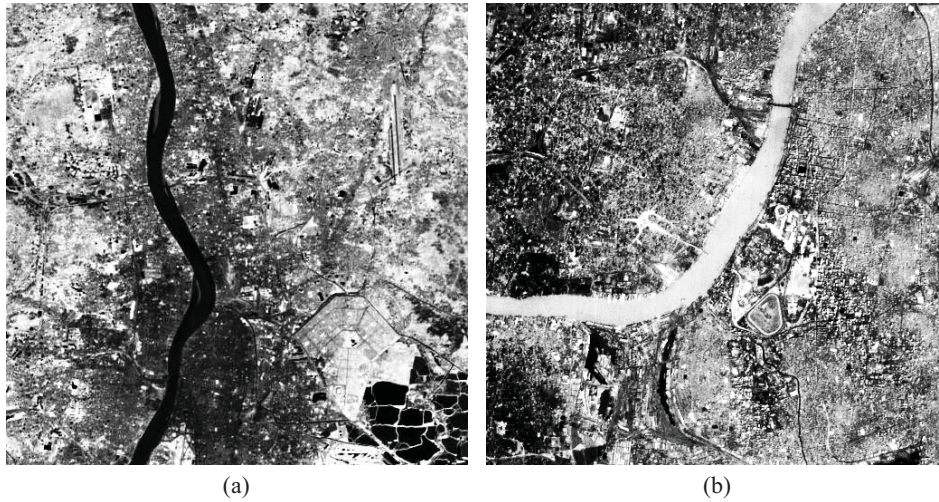
class with graded membership (Kuncheva, 2000). Many attempts have been made for remote sensing image analysis and classification using fuzzy sets (Chen, 1999/2000; Ghosh et al., 2008; Maselli et al., 1996; Melgani et al., 2000; Pal et al., 2000; Tso and Mather, 2001; Wang, 1990). It may be noted here that some attempts are also made to apply MCSs' (including neural-network as a combiner) to remote sensing image analysis (e.g., Giacinto and Roli, 1997; Smits, 2002; Steele, 2000; Debeir et al., 2002; Bruzzone and Cossu, 2002; Giacinto et al., 2000; Mahmoud et al., 2004; El-Melegy and Ahmed, 2007; Shankar et al., 2006, etc.)

In this scenario, the proposed technique has also been evaluated using a set of remote sensing images. Classification of multispectral remote sensing images are carried out using a set of training samples picked-up from the known land cover regions with the help of ground truth (as mentioned in Section 4.2.2). Using these labelled samples parameters are estimated for the classifiers. After learning, the classifier is used for classifying the land covers of the whole image. Results are presented on two images taken from two different satellites with different spatial and spectral resolutions.

4.2.1 Description of images

Out of the two multispectral images one is from Indian Remote Sensing-1A (IRS-1A) satellite (NRSA, 1989) and the other is from Systeme Pour d'Observation de la Terre (SPOT) satellite (Richards and Jia, 1999). These two images have different spatial and spectral resolutions.

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



We have used the IRS-1A image taken from the linear imaging self scanner (LISS-II) with spatial resolution of $36.25 \text{ m} \times 36.25 \text{ m}$ and wavelength range of $0.45\text{--}0.86 \mu\text{m}$. 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. Figure 2(a) shows an example of the enhanced band 4 image. The image in Figure 2(a) 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 belongs to TW class. CA class consists of buildings, runways, roads, bridges etc. Suburban and rural habitations, where concrete structures are comparatively less come under HAB class. VEG class represents crop and forest areas. OS class contains the barren land.

The SPOT multispectral image shown in Figure 2(b) is obtained from SPOT satellite (Richards and Jia, 1999). The Calcutta image used here has been acquired in the wavelength range of 0.50–0.89 μm . The whole spectrum range is decomposed into three spectral bands namely, green (band1), red (band2) and near infrared (band3) of wavelengths 0.50–0.59 μm , 0.61–0.68 μm , and 0.79–0.89 μm , respectively. This image has a spatial resolution of 20 m \times 20 m. We have considered the same six classes as in the case of IRS-1A image.

4.2.2 Performance evaluation indexes

The considered remote sensing images are partially labelled (i.e., only the training data has known labels, the test data is completely unlabeled). Using the ground truth information, the training data for each class was selected and all the classifiers along with combiners are trained on these labelled data. After training the classifiers and MCSs, the unlabeled data has been used for testing. Hence, we may call this datasets as partially labelled. Due to this limited information the performance measures used previously for the labelled datasets are not suitable for evaluation of the classification results. Thus, two clustering indexes namely, β index (Pal et al., 2000) based on hard labels and Xie-Beni (XB) index (Xie and Beni, 1991) based on soft labels are used. However, one can use other indexes like partition coefficient (PC), partition entropy (PE), Fukunyam-Sugeno (FS) index, and fuzzy hyper volume (FHV) (Wu and Yang, 2005) for evaluation of the results depending on the problem at hand. Wu and Yang (2005) analysed various fuzzy cluster validity indexes which includes PC, PE, FS and FHV. It is understood that PC, PE and FHV does not have any connection to the geometrical structures of the data, it only measures the compactness of the clusters. On the other hand, FS measures the error within the clusters and is not a good separation measure. Contrarily XB index addresses both the objectives, i.e., within cluster compactness and separation between the clusters. Due to this advantage and wide usability we have chosen the XB index in the present study.

β index

The β index (Pal et al., 2000) is defined 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 is expressed as:

$$\beta = \left(\sum_{i=1}^C \sum_{j=1}^{M_i} (x_{ij} - \bar{x})^2 \right) / \left(\sum_{i=1}^C \sum_{j=1}^{M_i} (x_{ij} - \bar{x}_i)^2 \right), \quad (13)$$

where \bar{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{x}_i is the mean of M_i pixel values of the i th class and $M = \sum_{i=1}^C M_i$, the total number of data points.

XB index

The XB index (Xie and Beni, 1991) provides a validity criterion based on a function that identifies overall compactness and separation of partition without any assumption about the number of substructures inherent in the data and is based on fuzzy labels. It is mathematically expressed in equation (14), as the ratio of compactness and separation.

$$XB = \frac{1}{M} \frac{\sum_{c=1}^C \sum_{i=1}^M \mu_{ci}^2 \|V_c - x_i\|^2}{\min_{c \neq j} \|V_c - V_j\|^2}, \quad (14)$$

where V_c and V_j , $\forall c, j=1, \dots, C$, are the centroids of the c th and j th class, respectively, and x_i is the i th pattern in the dataset. M is the total number of data points in the dataset and μ_{ci} is the membership value of the i th pattern to c th class. *The smaller the XB value, the better is the classification.*

4.2.3 Analysis of results

At first individual accuracies of all fuzzy classifiers are calculated using β and XB indexes and are depicted in Table 4. It can be observed that among the four classifiers the FPARR based method is providing the best results.

Table 4 β and XB values of different classification methods

Classification method	Image			
	IRS-1A		SPOT	
	β	XB	β	XB
Fk-NN	7.0121	0.9594	6.9212	2.5004
FMLC	7.0523	0.9356	6.9896	2.4231
FE	7.1312	0.9112	7.0137	2.3031
FPARR	8.1717	0.8310	8.1078	2.1021

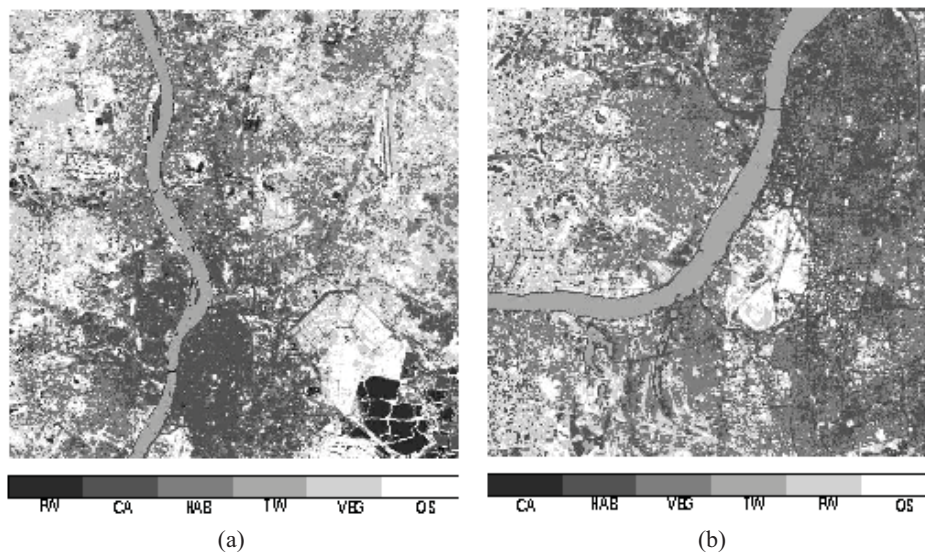
The validation results produced by MCS with different combination techniques are given in Table 5. From the table one can see that most of the MCSs with existing combiners are providing better results compared to any of the individual base classifiers. However, this improvement is not consistent for all images. For example, with IRS-1A image, better results are obtained for *DS* based MCS with β as the validity measure; whereas *fuzzy integral* based MCS showed improved performance with XB as the validity measure (Table 5). Results are completely different for SPOT image which provided better result 2 with *product* aggregation RR based MCS in terms of β validity measure, and *DT* based MCS in terms of XB validity measure. Hence, it is difficult to choose any of these six combiners for a particular image. The performance of the proposed NFC based MCS is then evaluated. The results revealed that the performance is further improved

consistently for both the images with respect to both the validity measures used here (Table 5). The classified images with this method are shown in Figure 3(a) and Figure 3(b). It can be seen from these images that all the land cover 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. This justified the consistency and superiority of the proposed NFC based MCS over all the six existing combiners for the considered datasets.

Table 5 β and XB values for different combination techniques

Combination technique	β index		XB index		
	IRS-1A image	SPOT image	IRS-1A image	SPOT image	
Voting	8.3134	8.2314	0.8211	2.1005	
Fuzzy aggregation reasoning rule	<i>Maximum</i>	8.2787	0.7903	2.1000	
	<i>Minimum</i>	8.3213	8.5134	0.7879	1.9733
	<i>Product</i>	8.6217	8.6321	0.8003	2.0178
	<i>Sum</i>	8.4312	8.3781	0.8202	2.0013
	<i>Mean</i>	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	
NFC	8.8012	8.7763	0.7697	1.8738	

Figure 3 Classified (a) IRS-1A image and (b) SPOT image using proposed NF combiner



5 Conclusions

A NFC based MCS for fusing the output of a set of fuzzy classifiers is proposed in the present article. The usefulness of it is demonstrated by showing its consistency in performance over other existing combiners for various datasets. Output of fuzzy classifiers are fed as input to a NN that acts as a learnable combiner. In the present study, we have considered a set of four base fuzzy classifiers. The outputs of these classifiers are combined in a multiple classifier paradigm with different combiners (existing and proposed NFC) and a performance analysis among them has been made. When we compared the performance of the proposed NFC based MCS with other similar existing combiner based MCSs, we observed that for the five standard datasets and two multispectral remote sensing images, the proposed model worked well and the improvement was consistent. On the other hand, results are different for different existing combiners and highly dependent on the input datasets. Thus, the superiority of the proposed technique is in its consistent behaviour over the existing combiners. In future, it may be interesting to compare the performance of the proposed NFC with other learnable combiners like support vector machine and Fisher discriminant analysis (the two successful classifiers).

Acknowledgements

The authors would like to thank the reviewers for their valuable suggestions. Also, they would like to thank the Department of Science and Technology, Government of India and the Ministry of Foreign Affairs, Government of Italy, and the Autonomous Province of Trento, Italy, the sponsors of the ITPAR program, under which a project titled *Advanced Techniques for Remote Sensing Image Processing* is being carried out at the Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India and the Department of Information and Communication Technologies, University of Trento, Trento, Italy.

References

- Alkoot, F.M. and Kittler, J. (1999) 'Experimental evaluation of expert fusion strategies', *Pattern Recognition Letters*, Vol. 20, Nos. 11–13, pp.1361–1369.
- Asuncion, A. and Newman, D.J. (2007) 'UCI machine learning repository', School of Information and Computer Sciences, University of California, Irvine, available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- Battiti, R. and Colla, A.M. (1994) 'Democracy in neural nets: voting schemes for classification', *Neural Networks*, Vol. 7, No. 4, pp.691–707.
- Bertolami, R. and Bunke, H. (2005) 'Ensemble methods for handwritten text line recognition systems', in *Proceedings of the 2005 IEEE International Conference on Systems, Man and Cybernetics*, pp.2334–2339.
- Bruzzone, L. and Cossu, R. (2002) 'A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps', *IEEE Trans. Geoscience and Remote Sensing*, Vol. 40, No. 9, pp.1984–1996.
- Bunke, H. and Kandel, A. (Eds.) (2000) *Neuro-Fuzzy Pattern Recognition*, World Scientific.
- Chen, C.F. (1999/2000) 'Fuzzy training data for fuzzy supervised classification of remotely sensed images', in *20th Asian Conference on Remote Sensing (ACRS 1999)*, pp.460–465.

- Cho, S.B. and Kim, J.H. (1995a) 'Combining multiple neural networks by fuzzy integral and robust classification', *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 25, pp.380–384.
- Cho, S.B. and Kim, J.H. (1995b) 'Multiple network fusion using fuzzy logic', *IEEE Transactions on Neural Networks*, Vol. 6, No. 2, pp.497–501.
- Debeir, O., Van Den Steen, I., Latinne, P., Van Ham, P. and Wolff, E. (2002) 'Textural and contextual land-cover classification using single and multiple classifier systems', *Photogrammetric Engineering and Remote Sensing*, Vol. 68, No. 6, pp.597–605.
- Demir, C. and Alpaydin, E. (2005) 'Cost-conscious classifier ensembles', *Pattern Recognition Letters*, Vol. 26, No. 14, pp.2206–2214.
- Drucker, H., Cortes, C., Jackel, L.D., LeCun, Y. and Vapnik, V. (1994) 'Boosting and other ensemble methods', *Neural Computation*, Vol. 6, pp.1289–1301.
- Elena database, available at <http://www.dice.ucl.ac.be/mlg/?page=Elena>.
- El-Melegy, M.T. and Ahmed, S.M. (2007) 'Neural networks in multiple classifier systems for remote-sensing image classification', in Nachtegael, M., Vin der Weken, D., Kerre, E.E. and Philips, W. (Eds.): *Soft Computing in Image Processing: Recent Advances, Volume 210 of Studies in Fuzziness and Soft Computing*, pp.65–96, Springer, Germany.
- Fumera, G. and Roli, F. (2005) 'A theoretical and experimental analysis of linear combiners for multiple classifier systems', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 6, pp.942–956.
- Gader, P.D., Mohamed, M.A. and Keller, J.M. (1996) 'Fusion of handwritten word classifiers', *Pattern Recognition Letters*, Vol. 17, No. 6, pp.577–584.
- Ghosh, A., Meher, S.K. and Shankar, B.U. (2008) 'A novel fuzzy classifier based on product aggregation operator', *Pattern Recognition*, Vol. 41, No. 3, pp.961–971.
- Giacinto, G. and Roli, F. (1997) 'Ensembles of neural networks for soft classification of remote sensing images', in *Proc. of the European Symposium on Intelligent Techniques*, pp.166–170, Italy.
- Giacinto, G., Roli, F. and Bruzzone, L. (2000) 'Combination of neural and statistical algorithms for supervised classification of remote-sensing images', *Pattern Recognition Letters*, Vol. 21, No. 5, pp.385–397.
- Gutta, S., Huang, J.R.J., Jonathon, P. and Wechsler, H. (2000) 'Mixture of experts for classification of gender, ethnic origin, and pose of human faces', *IEEE Transactions on Neural Networks*, Vol. 11, No. 4, pp.948–960.
- Hansen, L.K. and Salamon, P. (1990) 'Neural network ensembles', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 10, pp.993–1001.
- Hashem, S. (1997) 'Optimal linear combinations of neural networks', *Neural Networks*, Vol. 10, No. 4, pp.599–614.
- Hashem, S. and Schmeiser, B. (1995) 'Improving model accuracy using optimal linear combinations of trained neural networks', *IEEE Transactions on Neural Networks*, Vol. 6, No. 3, pp.792–794.
- Haykin, S. (1998) *Neural Networks: A Comprehensive Foundation*, 2nd ed., Prentice Hall.
- Ho, T.K., Hull, J.J. and Srihari, S.N. (1994) 'Decision combination in multiple classifier systems', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 1, pp.66–75.
- Huang, Y.S. and Suen, C.Y. (1994) 'A method of combining multiple classifiers: a neural network approach', in *International Conference on Pattern Recognition*, Vol. 2, pp.473–475.
- Huang, Y.S., Liu, K. and Suen, C.Y. (1995) 'The combination of multiple classifiers by a neural network approach', *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 9, No. 3, pp.579–597.
- International Workshop on Multiple Classifier Systems* (2000–2005 and 2007) LNCS, Vols. 1857, 2096, 2364, 2709, 3077, 3541 and 4472, Springer.
- Jacobs, R.A. (1995) 'Methods for combining experts' probability assessments', *Neural Computation*, Vol. 7, No. 5, pp.867–888.

- Jain, A.K., Duin, R.P.W. and Mao, J. (2000) 'Statistical pattern recognition: a review', *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 22, No. 1, pp.4–37.
- Jang, J.S.R., Sun, C-T. and Mizutani, E. (1996) *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Pearson Education.
- Keller, J.M., Gray, M. and Givens, J. (1985) 'A fuzzy k-nearest neighbor algorithm', *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 15, No. 4, pp.580–585.
- Kimura, F. and Shridhar, M. (1991) 'Handwritten numerical recognition based on multiple algorithms', *Pattern Recognition*, Vol. 24, No. 10, pp.969–983.
- Kittler, J. (1997) 'Improving recognition rates by classifier combination: a review', in *Proc. First IAPR TCI Workshop Statistical Techniques in Pattern Recognition*, pp.205–210.
- Kittler, J. (1998) 'Combining classifiers: a theoretical framework', *Pattern Analysis and Applications*, Vol. 1, pp.18–27.
- Kittler, J. (2002) 'Multiple classifier system', in Ghosh, A. and Pal, S.K. (Eds.): *Soft Computing Approach to Pattern Recognition and Image Processing, Volume 53 of Machine Perception and Artificial Intelligence*, pp.3–22, World Scientific.
- Kittler, J., Hatef, M., Duin, R.P.W. and Matas, J. (1998) 'On combining classifiers', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, pp.226–239.
- Krogh, A. and Vedelsby, J. (1995) 'Neural network ensembles, cross validation, and active learning', in Tesauro, G., Touretzky, D. and Leen, T. (Eds.): *Advances in Neural Information Processing Systems*, Vol. 7, pp.231–238, The MIT Press, Cambridge MA.
- Kuncheva, L.I. (2003) *Combining Classifiers: Soft Computing Solutions*, World Scientific, Chapter 15, pp.427–451.
- Kuncheva, L.I. (2000) *Fuzzy Classifier Design*, Springer-Verlag.
- Kuncheva, L.I. (2001) 'Using measures of similarity and inclusion for multiple classifier fusion by decision templates', *Fuzzy Sets and Systems*, Vol. 122, No. 3, pp.401–407.
- Kuncheva, L.I. (2002a) 'A theoretical study on six classifier fusion strategies', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, pp.281–286.
- Kuncheva, L.I. (2002b) 'Switching between selection and fusion in combining classifiers: an experiment', *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, Vol. 32, No. 2, pp.146–156.
- Kuncheva, L.I. (2004) *Combining Pattern Classifiers: Methods and Algorithms*, Wiley-Interscience.
- Kuncheva, L.I., Bezdek, J.C. and Duin, R.P.W. (2001) 'Decision templates for multiple classifier fusion: an experimental comparison', *Pattern Recognition*, Vol. 34, pp.299–314.
- Lee, D-S. (1995) 'A theory of classifier combination: the neural network approach', PhD dissertation, State University of New York, Buffalo.
- Lee, D-S. and Srihari, S.N. (1995a) 'A theory of classifier combination: the neural network approach', in *ICDAR '95: Proceedings of the Third International Conference on Document Analysis and Recognition*, Vol. 1, pp.42–45, IEEE Computer Society, Washington, DC, USA.
- Lee, D-S. and Srihari, S.N. (1995b) 'Dynamic classifier combination using neural network', in *Proc. SPIE – Doc. Rec. II*, pp.26–37.
- Lin, C.T. and Lee, C.S.G. (1996) *Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems*, Prentice Hall, New Jersey.
- Liu, K., Huang, Y.S. and Suen, C.Y. (1994) 'Image classification by classifier combining technique', in Chen, S-S. (Ed.): *Proc. SPIE Neural and Stochastic Methods in Image and Signal Processing III*, Vol. 2304, pp.210–217.
- Mahmoud, S., El-Melegy, M.T. and Farag, A.A. (2004) 'A comparative study of statistical and neural methods for remote sensing image classification and decision fusion', in *International Conference on Image Processing (ICIP'04)*, pp.3347–3350.

- Maselli, F., Rodolfi, A. and Copnese, C. (1996) 'Fuzzy classification of spatially degraded thematic mapper data for the estimation of sub-pixel components', *International Journal of Remote Sensing*, Vol. 17, No. 3, pp.537–551.
- Meher, S.K., Ghosh, A., Shankar, B.U. and Bruzzone, L. (2006) 'Neuro-fuzzy fusion: a new approach to multiple classifier system', in *Proceedings 9th International Conference on Information Technology*, pp.209–212, IEEE Computer Society, Los Alamitos, CA, USA.
- Melgani, F., Al Hashemy, B.A.R. and Taha, S.M.R. (2000) 'An explicit fuzzy supervised classification method for multispectral remote sensing images', *IEEE Transaction on Geoscience and Remote Sensing*, Vol. 38, No. 1, pp.287–295.
- Mitra, S. and Hayashi, Y. (2000) 'Neuro-fuzzy rule generation: survey in soft computing framework', *IEEE Transactions on Neural Networks*, Vol. 11, No. 3, pp.748–768.
- NRSA (1989) 'IRS data users hand book', Technical report, NRSA, Document No. IRS/NRSA/NDC/HB-02/89.
- Pal, S.K. and Majumder, D.D. (1977) 'Fuzzy sets and decision making approaches in vowel and speaker recognition', *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 7, No. 1, pp.625–629.
- Pal, S.K. and Mitra, S. (1999) *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing*, John Wiley & Sons, Inc., NY, USA.
- Pal, S.K., Ghosh, A. and Shankar, B.U. (2000) 'Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation', *International Journal of Remote Sensing*, Vol. 21, No. 11, pp.2269–2300.
- Polikar, R. (2006) 'Ensemble based systems in decision making', *IEEE Circuits and Systems Magazine*, Vol. 6, No. 3, pp.21–45.
- Raudys, S. (2006a) 'Trainable fusion rules: I. large sample size case', *Neural Networks*, Vol. 19, No. 10, pp.1506–1516.
- Raudys, S. (2006b) 'Trainable fusion rules: II. small sample-size effects', *Neural Networks*, Vol. 19, No. 10, pp.1517–1527.
- Richards, J.A. and Jia, X. (1999) *Remote Sensing Digital Image Analysis: An Introduction*, 3rd ed., Springer, New York.
- Rogova, G. (1994) 'Combining the results of several neural network classifiers', *Neural Networks*, Vol. 7, No. 5, pp.777–781.
- Shankar, B.U., Meher, S.K., Ghosh, A. and Bruzzone, L. (2006) 'Remote sensing image classification: a neuro-fuzzy MCS approach', in Kalra, P.K. and Peleg, S. (Eds.): *Proceedings 5th Indian Conference Computer Vision, Graphics and Image Processing*, pp.128–139.
- Smits, P.C. (2002) 'Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection', *IEEE Transaction on Geoscience and Remote Sensing*, Vol. 40, No. 4, pp.801–813.
- Steele, B.M. (2000) 'Combining multiple classifiers. An application using spatial and remotely sensed information for land cover type mapping', *Remote Sensing of Environment*, Vol. 74, No. 3, pp.545–556.
- Tahani, H. and Keller, J.M. (1990) 'Information fusion in computer vision using the fuzzy integral', *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 20, No. 3, pp.733–741.
- Tax, D., van Breukelen, M., Duin, R. and Kittler, J. (2000) 'Combining multiple classifiers by averaging or by multiplying', *Pattern Recognition*, Vol. 33, No. 9, pp.1475–1485.
- Theodoridis, S. and Koutroumbas, K. (2006) *Pattern Recognition*, 3rd ed., Academic Press, USA.
- Ting, K.M. and Witten, I.H. (1999) 'Issues in stacked generalization', *Journal of Artificial Intelligence Research*, Vol. 10, pp.271–289.
- Tso, B. and Mather, P.M. (2001) *Classification Methods for Remotely Sensed Data*, Taylor and Francis, London.
- Wang, F. (1990) 'Fuzzy supervised classification of remote sensing images', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 28, pp.194–201.

- Witten, I.H. and Frank, E. (2005) *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed., Morgan Kaufmann, USA.
- Wolpert, D.H. (1992) 'Stacked generalization', *Neural Networks*, Vol. 5, No. 2, pp.241–260.
- Wu, K-L. and Yang, M-S. (2005) 'A cluster validity index for fuzzy clustering', *Pattern Recognition Letters*, Vol. 26, No. 9, pp.1275–1291.
- Xie, X.L. and Beni, G. (1991) 'A validity measure for fuzzy clustering', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 8, pp.841–847.
- Xu, L., Krzyzak, A. and Suen, C.Y. (1992) 'Methods of combining multiple classifiers and their applications to handwriting recognition', *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22, No. 3, pp.418–435.
- Zadeh, L.A. (1965) 'Fuzzy sets', *Information Control*, Vol. 8, pp.338–353.