Neural Networks 22 (2009) 100–109

A novel approach to neuro-fuzzy classification

Ashish Ghosh *, B. Uma Shankar, Saroj K. Meher

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

ARTICLE INFO

Article history:
Received 17 November 2006
Revised and accepted 18 September 2008

Keywords:
Pattern classification
Fuzzy logic
Neural networks
Neuro-fuzzy

ABSTRACT

A new model for neuro-fuzzy (NF) classification systems is proposed. The motivation is to utilize the feature-wise degree of belonging of patterns to all classes that are obtained through a fuzzification process. A fuzzification process generates a membership matrix having total number of elements equal to the product of the number of features and classes present in the data set. These matrix elements are the input to neural networks. The effectiveness of the proposed model is established with four benchmark data sets (completely labeled) and two remote sensing images (partially labeled). Different performance measures such as misclassification, classification accuracy and kappa index of agreement for completely labeled data sets, and β index of homogeneity and Davies–Bouldin (DB) index of compactness for remotely sensed images are used for quantitative analysis of results. All these measures supported the superiority of the proposed NF classification model. The proposed model learns well even with a lower percentage of training data that makes the system fast.

1. Introduction

Many attempts have been made in the last decades to design hybrid systems for pattern classification by combining the merits of individual techniques. An integration of neural networks (NNs) and fuzzy systems is one such hybrid technique and is known as neuro-fuzzy (NF) computing (Abe, 2001; Pal & Ghosh, 1996; Pal & Mitra, 1999).

Uncertainties can arise at any stage of a pattern classification system, resulting from incomplete or imprecise input information, ambiguity or vagueness in input data, ill-defined and/or overlapping boundaries among classes or regions, and indefiniteness in defining/extracting features and relations among them. It is therefore necessary for a classification system to have sufficient provision for representing uncertainties involved at every stage so that the final output (results) of the system is associated with the least possible uncertainty. The uncertainty handling issue becomes more prominent in case of land cover classification of remote sensing imagery (Richards & Jia, 2006; Tso & Mather, 2001; Varshney & Arora, 2004).

Since the fuzzy set theory (Zadeh, 1965) is a generalization of the classical set theory, it has greater flexibility to capture various aspects of incompleteness or imperfection about real life situations. The significance of fuzzy set theory in the realm of pattern classification is effectively justified in various areas such as representing input patterns as an array of membership values denoting the degree of possession of certain properties, representing linguistically defined input features, representing multiclass membership of ambiguous patterns, generating rules and inferences in linguistic form, extracting ill-defined image regions, and describing relations among them (Pal, Ghosh, & Kundu, 2000; Pedrycz, 1990).

Neural networks (NNs) are aimed at emulating the biological nervous system with the hope of achieving human-like performance artificially by capturing the key ingredients responsible for the remarkable capabilities of the human nervous system (Anthony & Bartlett, 1999; Haykin, 1997; Ripley, 1996; Rumelhart, Hinton, & Williams, 1986). Interaction among the neurons is very high in NNs making them suitable for collective decision making. The main characteristics of NNs, namely, adaptivity, fault tolerance, robustness and optimality play important roles particularly in the field of pattern classification.

Both NNs and fuzzy systems are adaptive in the estimation of the input–output function without any precise mathematical model. NNs handle numeric and quantitative information while fuzzy systems can handle symbolic and qualitative data. Apart from this, in a fuzzy classifier patterns are assigned with a degree of belonging to different classes. Thus the partitions in fuzzy classifiers are soft and gradual rather than hard and crisp. Therefore, an integration of neural and fuzzy systems should have the merits of both and it should enable one to build more intelligent decision making systems. Fuzzy set theory based hybrid classification systems are found to be more suitable and appropriate to handle these situations reasonably (Kuncheva, 2000; Pedrycz, 1990).

* Corresponding author.
E-mail address: ash@isical.ac.in (A. Ghosh).

In the NF paradigm, much research effort has been made (Abe, 2001; Baraldi, Binaghi, Blonda, Brivio, & Rampini, 2001; Boskovitz & Guterman, 2002; Gamba & Dellacqua, 2003; Ghosh, Pal, & Pal, 1993; Han, Lee, Chi, & Ryu, 2002; Keller & Hunt, 1985; Kwon, Ishibuchi, & Tanaka, 1994; Pal & Ghosh, 1996; Pal & Mitra, 1999; Qiu & Jensen, 2004). NF hybridization is done broadly in two ways: NNs that are capable of handling fuzzy information (named as *fuzzy-neural networks* (FNN)), and fuzzy systems augmented by NNs to enhance some of their characteristics such as flexibility, speed and adaptability (named as *neural-fuzzy systems* (NFS)) (Pal & Ghosh, 1996; Pal & Mitra, 1999). Other than these two, fuzzy sets/logic can also be incorporated in NNs in various ways. All these methodologies can be broadly categorized into five NF integration procedures, and the details on these methodologies can be found in Pal and Ghosh (1996).

The main aim of the present work is to explore various possible degrees of belonging of all features independently to different classes; normally not used in conventional NF classification systems. The proposed hybrid classification model assigns memberships for each feature of a pattern to different classes forming the *membership matrix*. The number of columns and rows of the matrix are equal to the number of classes and number of features (spectral bands for remote sensing images), respectively. Therefore, the input vector will have a dimension equal to the product of the number of classes and the number of features. In other words, the number of input nodes of the NN is equal to the number of elements of the *membership matrix*. This membership matrix is converted into a vector by cascading all rows (columns) and becomes the input to the NN. Number of output nodes of the NN is equal to the number of classes. Defuzzification operation is then performed on the NN output. A hard classification of the input pattern can be obtained using a *MAX* (maximum) operation on the output of NN as in the case of a conventional fuzzy classification system.

The organization of rest of the article is as follows. A detailed description of the proposed NF classification model has been made in Section 2. Section 3 describes the experimental results with comparative analysis. Various performance measures used in the present investigation are also discussed in this section. Finally, concluding remarks are provided in Section 4.

2. Proposed neuro-fuzzy classification model

A new model for the neuro-fuzzy (NF) classification system is proposed in the present article. The proposed NF classification system extracts feature-wise information of input pattern to different classes. Since all features are not equally important in discriminating all classes, the feature-wise belonging is expected to help in the classification process. The block diagram of the proposed NF model is shown in Fig. 1.

The proposed model works in three steps. In the first step, the system takes an input and fuzzifies its feature values using membership functions (MF), and provides the membership of individual features to different classes. A *membership matrix* thus formed contains number of rows and columns equal to the number of features and classes, respectively, present in a data set. In the present study, we have used a popular π-type MF to model a class (Pal & Majumder, 1977). Thus, the first step of the proposed NF classification system extracts the hidden or interrelated information of features to all classes through the MF that may be helpful for obtaining better classification accuracy. The advantage of using π-type MF is that it has a parameter, called fuzzifier (m), which can be tuned easily according to the requirement of the problem. This provides more flexibility for classification. Thus the generalization capability can be controlled by selecting a proper value of m.

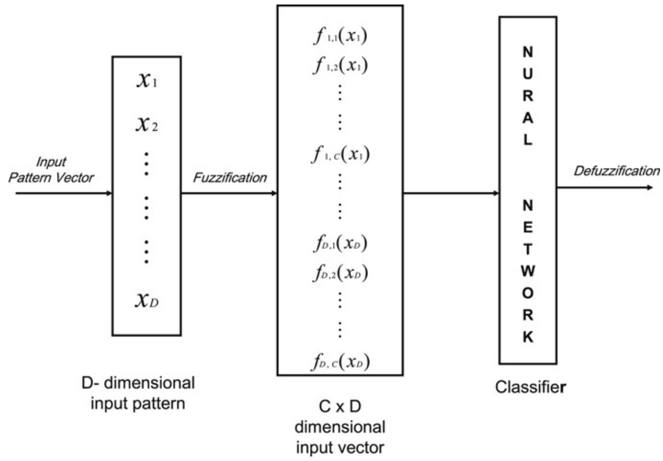

Fig. 1. Proposed neuro-fuzzy classification model.

In the second step, the *membership matrix* is converted into a vector by cascading all rows or columns. This vector becomes the input to the NN and thus the number of input nodes of the NN is equal to the product of the number of features and classes. The number of output nodes in the NN is the same as the number of classes present in the data set.

The last step of the proposed NF classifier is a hard classification by performing a *MAX* operation to defuzzify the output of the NN. A pattern is assigned to class c with the highest class membership value. However, we can also use the fuzzy output of NN for higher level processing, if desired, particularly for image analysis problems.

2.1. Fuzzification

The MF generates a feature-wise degree of belonging of a pattern to different classes by fuzzification. The membership matrix $f_{d,c}(x_d)$ thus generated, expresses the degree of belonging of different features (D) to different classes (C), where x_d is the dth feature value of pattern $\mathbf{x}$; with $d = 1, 2, \ldots, D$ and $c = 1, 2, \ldots, C$. A pattern is thus represented as

$$\mathbf{x} = [x_1, x_2, \ldots, x_d, \ldots, x_D]^{\mathrm{T}}. \tag{1}$$

Various types of MFs are used in fuzzy systems for modeling input values. Here we have used a popular π-type MF to model a class (Ghosh, Meher, & Shankar, 2008; Pal & Majumder, 1986). Note that it is a bounded function having a shape similar to that of Gaussian/exponenital function; and by varying the value of the fuzzifier m we can control the steepness of the function. The function is defined as (shown in Fig. 2):

$$\begin{aligned}
\pi(x; a, r, b) &= 0, && x \le a \\
&= 2^{m-1}[(x-a)/(r-a)]^m, && a < x \le p \\
&= 1 - 2^{m-1}[(r-x)/(r-a)]^m, && p < x \le r \\
&= 2^{m-1}[(x-r)/(b-r)]^m, && r < x \le q \\
&= 1 - 2^{m-1}[(b-x)/(b-r)]^m, && q < x < b \\
&= 0, && x \ge b
\end{aligned} \tag{2}$$

where m is called the fuzzifier. In the present investigation, we have selected the fuzzifier value as 2. The MF has a center at r, with $r = (p + q)/2$, where p and q are the two crossover points. Membership value at the crossover points is 0.5 and at the center r its value is 1.0 (maximum). Assignment of membership value is made in such a way that a training data gets a membership value of 1.0 when its feature value is at the center of the MF, and when it is away from the center its value gradually decreases and attains 0.5 at the boundary of the training set. The extended

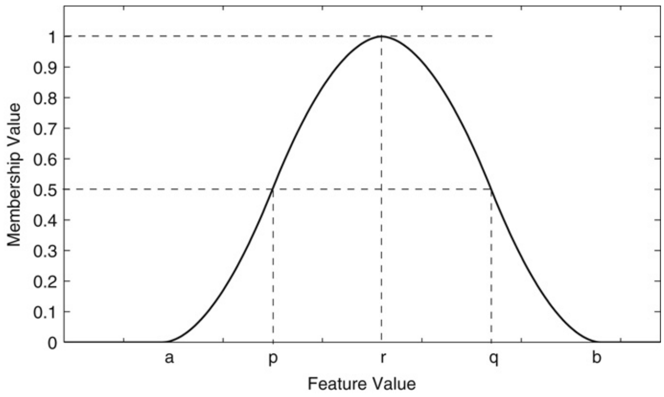

Fig. 2. π-type membership function used to compute the membership value of a feature to a class.

region beyond p and q will have membership values less than 0.5. This region is considered to incorporate the portion (of the pattern class) possibly uncovered by the training samples, which provides a scope/flexibility for improved generalization ability of the system. The center r is computed as the *mean* of the training data set. It is defined as $r = mean(y)$ (i.e., average value of the data set for a feature y). The crossover points p and q are estimated as $p = mean(y) - [max(y) - min(y)]/2$, and $q = mean(y) + [max(y) - min(y)]/2$, where *min* and *max* are the minimum and maximum value, respectively, of the data set for a particular feature y. Such a choice of p and q ensure that most of the training patterns have a membership values $\geq$0.5; and test patterns can have membership value in [0, 1]. For a pattern $\mathbf{x}$, the membership matrix after fuzzification process is expressed as:

$$F(\mathbf{x}) = \begin{bmatrix} f_{1,1}(x_1) & f_{1,2}(x_1) & \cdots & f_{1,C}(x_1) \\ f_{2,1}(x_2) & f_{2,2}(x_2) & \cdots & f_{2,C}(x_2) \\ \cdots & \cdots & \cdots & \cdots \\ f_{D,1}(x_D) & f_{D,2}(x_D) & \cdots & f_{D,C}(x_D) \end{bmatrix} \quad (3)$$

where $f_{d,c}(x_d)$ represents the membership of the dth feature to the cth class.

For example, $f_{2,3}(x_2)$ represents the membership grade of the 2nd feature to class 3. This fuzzified pattern matrix (membership matrix) is used as input to a NN as described below.

2.2. Neural networks

The proposed NF classification method has been implemented using the most popular feed forward multi-layer perceptron classifier (Anthony & Bartlett, 1999; Baldi & Homik, 1995; Haykin, 1997; Marinai, Gori, & Soda, 2005; Ripley, 1996; Zhang, 2000; Zurada, 1992) having three layers — known as input, hidden and output layers, respectively. The number of nodes in the input-layer is equal to the number of elements in the membership matrix and the number of nodes in the output-layer is equals to the number of classes present in the data set. Number of nodes in the hidden layer is chosen to be equal to the square root of the product of the number nodes in the input and output layers (Rumelhart et al., 1986). We have used a single hidden-layer for the present investigation.

2.3. Defuzzification

The last step of the proposed NF system is a hard classification by performing a *MAX* (maximum) operation to defuzzify the output of the NN. The pattern is assigned to class c corresponding to the highest output value. Mathematically, assign the pattern to class c if

$$F_c(\mathbf{x}) \geq F_j(\mathbf{x}) \quad \forall j \in 1, 2, \ldots, C \text{ and } j \neq c \quad (4)$$

where $F_j(\mathbf{x})$ is the activation value of the jth neuron in the output layer.

However, the output of the NF system can be used as fuzzy output also for further analysis, if desired.

3. Experimental results and analysis

For establishing the usefulness of the proposed model we considered four conventional fully labeled data sets (including a satellite image data) and two partially labeled multi-spectral remote sensing images. A brief description of the conventional (labeled) data sets is given in Table 1, and the remote sensing images used are shown in Figs. 3(a) and 5(a).

Selection of the training and test samples for all classes in case of conventional (fully labeled) data sets have been made after dividing the whole data set into two parts. The first part (training data) is taken for estimation of the parameters of the classifiers. The second part (test data) is taken for testing the performance. We have taken 10%, 20% and 50% as training data and the rest 90%, 80% and 50% are considered as test data. Selection of the training data is random and an equal percent of data is collected from each class. This means 10% or 20% or 50% of data from all available classes of the data sets have been used for training purpose. However the selection of the training samples for all classes in case of multispectral remote sensing images (partially labeled) is made according to a prior assumption of the land cover regions. A brief description of the quantitative measures used for the evaluation of the proposed method are provided in the following sections.

3.1. Performance measurement indexes

To examine the practical applicability of the proposed classification model we used *Misclassification (MC), Percentage of overall class Accuracy (PA)* and *Kappa Index of Agreement (KIA)* (Card, 1982; Congalton, 1991) as performance measures (for fully labeled data sets). The MC value denotes the number of overall sample patterns that are wrongly classified. The PA value shows the total percentage of correctly classified sample patterns. The MC and PA parameters are calculated with respect to the total number of patterns with true class labels. They do not provide the class-wise agreement/matching between the true and estimated class labels. Thus, to get an overall class-wise agreement based on the individual class accuracy, we have used KIA. A good KIA value signifies better agreement of the estimated data with the true one. The KIA value is estimated from a confusion or error matrix (CM) (Card, 1982; Congalton, 1991). A CM is a square assortment of numbers defined in rows and columns that represents the number of sample patterns assigned to a particular class relative to the true class. This matrix produces many statistical measures of class accuracy including overall classification accuracy (the sum of the diagonal elements divided by the total number of samples) and KIA. KIA is defined as:

$$KIA = \frac{N \sum_{i=1}^{r} X_{ii} - \sum_{i=1}^{r} (X_{i+}.X_{+i})}{N^2 - \sum_{i=1}^{r} (X_{i+}.X_{+i})} \quad (5)$$

where the confusion matrix (CM) has

$$\begin{aligned} r &= \text{number of rows}, \\ X_{ii} &= \text{number of observations in row } i \text{ and column } i, \\ X_{i+} &= \text{total number of observation in row } i, \\ X_{+i} &= \text{total number of observation in column } i, \text{ and} \\ N &= \text{total number of observations in the CM}. \end{aligned} \quad (6)$$

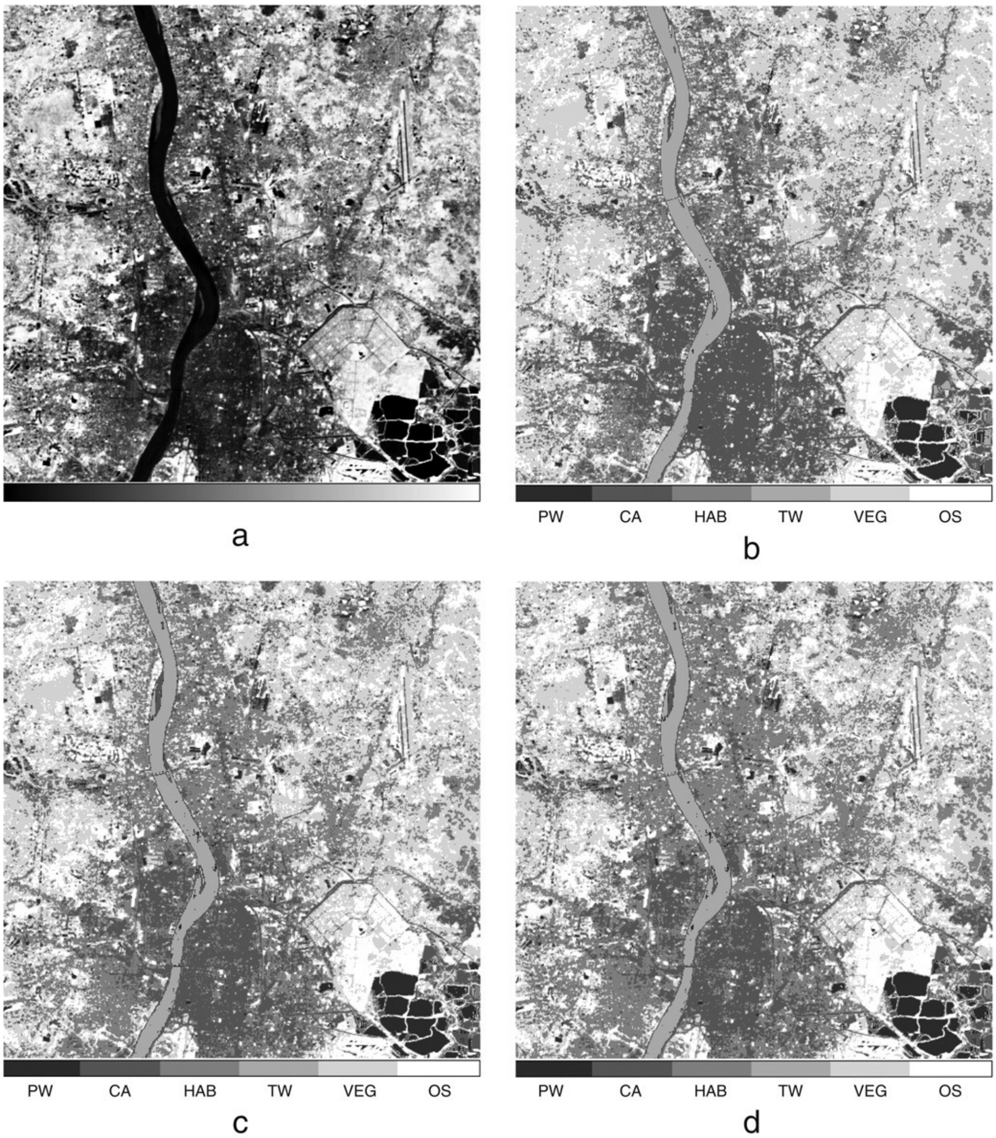

Fig. 3. IRS-1A Calcutta images: (a) Enhanced input, (b) land cover classified by MLP, (c) land cover classified by conventional NF, and (d) land cover classified by proposed NF methods.

Table 1
Summary of the benchmark (fully labeled) data sets used in the present study.

Name of the data set	Number of classes	# of features available	# of features used	# of patterns
VOWEL	6	3	3	871
PHONEME	2	5	5	5404
BLOCKS	5	10	10	5473
SATIMAGE	6	36	4[a]	6435

[a] As suggested by the contributor of the data set.

For the remote sensing image data, a very small set of training patterns is picked up from the known regions for classifying rest of the image (unlabeled). Therefore, it is not possible to assess the results with the indexes described above. Hence, we have used two other indexes, one is β index (Pal, Ghosh, & Shankar, 2000) of *homogeneity* and the other one is Davies–Bouldin (*DB*) index (Davies & Bouldin, 1979) of *compactness and separability* and they are discussed below.

β index of homogeneity

The β index has been successfully used in the assessment of image segmentation quality (Acharyya, De, & Kundu, 2003; Mitra, Shankar, & Pal, 2004; Pal et al., 2000). It is defined as the ratio of the total variation and within-class variation (Pal et al., 2000). Since the numerator is constant for a given image, β value is dependent only on the denominator. *The denominator decreases with increase in homogeneity within the class for a fixed number of classes* (C). Thus for a given image and a given number of classes (land covers), *the higher the homogeneity within the classes, the higher would be the β value*. Mathematically β is represented as:

$$\beta = \frac{\sum_{i=1}^{C}\sum_{j=1}^{M_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}})^2}{\sum_{i=1}^{C}\sum_{j=1}^{M_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^2} \tag{7}$$

where $\bar{\mathbf{x}}$ is the mean grey value of the image pixels (pattern vector), M_i is the number of pixels in the ith ($i = 1, 2, \ldots, C$) class, $\mathbf{x}_{ij}$ is the

Table 2
Misclassification (MC), Percentage of overall class Accuracy (PA) and *Kappa Index of Agreement (KIA)* values for VOWEL data set.

Classification method	% training data								
	10			20			50		
	MC	PA	KIA	MC	PA	KIA	MC	PA	KIA
MLP	179	77.08	0.7123	153	77.95	0.7210	92	78.80	0.7411
Conventional NF	166	78.74	0.7412	143	79.39	0.7740	86	80.18	0.7835
Proposed NF	**148**	**81.04**	**0.8001**	**128**	**81.55**	**0.8123**	**79**	**81.79**	**0.8214**

grey value of the jth pixel ($j = 1, 2, \ldots, M_i$) in class i, and $\bar{\mathbf{x}}_i$ is the mean of M_i grey values of the ith class.

Further, in the present work we have evaluated the corresponding percentage of gain with β values $Gain_\beta$ using the following formula:

$$Gain_\beta = \frac{(\beta \text{ value with classifier-2} - \beta \text{ value with classifier-1})}{\beta \text{ value with classifier-1}} \times 100 \tag{8}$$

where classifier-1 is the old classifier and classifier-2 is a relatively new classifier (the classifier for which the gain is measured) which is performing better than the old one.

Davies–Bouldin (DB) index of compactness and separability

Davies–Bouldin *(*DB) index for cluster validation has been defined and used in Davies and Bouldin (1979). Here we are using the index for validating our classification results on partially labeled data sets. The idea behind *DB* index is that, for a good partition, inter-class separation as well as intra-class homogeneity and compactness should be high. The *DB* index is based on the evaluation of some measure of dispersion S_i within the ith cluster and the distance (d_{ij}) between the prototypes of clusters i and j. It is defined as:

$$S_{i,q} = \left(\frac{1}{|X_i|} \sum_{\mathbf{x} \varepsilon X_i} \|\mathbf{x} - \mathbf{v}_i\|_2^q \right)^{\frac{1}{q}}, \tag{9}$$

and

$$d_{ij,t} = \left[\sum_{s=1}^{p} |v_{si} - v_{sj}|^t \right]^{\frac{1}{t}}. \tag{10}$$

$S_{i,q}$ is the qth root of the qth moment of the points in cluster i with respect to their mean or centroid ($\mathbf{v}_i$), and is a measure of dispersion of the points ($\mathbf{x}$) in cluster i. $|X_i|$ is the cardinality of cluster i. $d_{ij,t}$ is the Minkowski's distance of order t between the centroids of the extracted clusters i and j. In the present experiment we have taken $q = t = 2$. We compute:

$$R_{i,qt} = \max_{j=1,2,\ldots,C \text{ and } j \neq i} \left[\frac{S_{i,q} + S_{j,q}}{d_{ij,t}} \right], \quad \forall i = 1, 2, \ldots, C. \tag{11}$$

The DB index is then defined as:

$$DB = \frac{1}{C} \sum_{i=1}^{C} R_{i,qt}, \tag{12}$$

with C as the number of clusters/classes. The smaller the *DB* value, better is the partitioning (Davies & Bouldin, 1979). The corresponding percentage of gain obtained with *DB* values (i.e., $Gain_{DB}$) is computed as:

$$Gain_{DB} = \frac{(DB \text{ value with classifier-1} - DB \text{ value with classifier-2})}{DB \text{ value with classifier-1}} \times 100. \tag{13}$$

3.2. *Classification of conventional data sets (i.e., completely labeled data sets)*

We have chosen four benchmark data sets for the present experiments. The data sets used here are VOWEL, PHONEME, BLOCKS and SATIMAGE. The BLOCKS and SATIMAGE data sets are chosen from machine learning repository at UCI, USA (Asuncion & Newman, 2007) and PHONEME data set is from ELENA Database at MLG/UCL, Belgium (ELENA Database, 1995), whereas the VOWEL data set is from our institute's collection (Pal & Majumder, 1977).

3.2.1. *VOWEL data*

The VOWEL data is a set of Indian Telegu vowel sounds in consonant–vowel–consonant context uttered by three speakers in the age group 30–35 years (Pal & Majumder, 1977). The data set consists of 871 sample patterns. It has three features and six classes /δ/, / a /, / i /, / u /, / e / and / o / with 72, 89, 172, 151, 207 and 180 samples, respectively. The classes are highly overlapping.

Table 2 depicts the results of classification of VOWEL data set for three different percentages of training sets. For 10% training data MLP provided an MC of 179 and the PA as 77.08. These values are 166 and 78.74 with conventional NF method. The proposed NF method provided an improved MC and PA values and they are 148 and 81.04. Thus there is an increase of nearly 4% of PA by the proposed method, and the corresponding MC value is decreased by 31 compared to MLP. Similarly an increase in PA of nearly 3% and decrease in MC of 18 is obtained compared with the conventional NF classification method. For 20% training data, nearly the same PA increment is there as in the case of 10%, i.e., with the proposed NF, conventional NF and MLP methods these values are 81.55, 79.39 and 77.95, respectively. Also there is a decrease in MC of 15 with the proposed NF method compared with the conventional NF and of 25 with MLP. A similar improvement with the proposed NF classification can be observed with 50% training data. Table 2 also reveals that the accuracy obtained with the proposed NF classifier for minimum percentage of training data is nearly the same as with the conventional NF and MLP at 50% training data. This is particularly important when there is a scarcity of training data set (e.g., in classification of remote sensing images).

The superiority of the proposed NF classification method is also validated with KIA as shown in Table 2. A comparison is made among the three methods mentioned earlier. Table 2 shows that the KIA with the proposed NF method at 10% training data is 0.8001, which is more than 0.7412 and 0.7123 obtained using conventional NF and MLP. A similar improvement of KIA value is obtained with 20 and 50 percent training data. For example, at 20% training data, the KIA values are 0.8123, 0.7740 and 0.7210, respectively. From these values, it is obvious that the proposed NF classification method provided a better class-wise accuracy along with overall accuracy compared with the conventional NF and MLP methods. It is true for all percentages of training data. The KIA value also justifies that at the minimum percentage of training data the classification accuracy (class-wise and overall) of the proposed NF method is more promising than conventional NF and MLP.

Table 3
Misclassification (MC), Percentage of overall class Accuracy (PA) and *Kappa Index of Agreement (KIA)* values for PHONEME data set.

Classification method	% training data								
	10			20			50		
	MC	PA	KIA	MC	PA	KIA	MC	PA	KIA
MLP	1245	74.39	0.4012	1079	75.03	0.4137	662	75.49	0.5651
Conventional NF	1160	76.14	0.4672	1009	76.65	0.4905	613	77.31	0.5873
Proposed NF	**968**	**80.09**	**0.5861**	**832**	**80.74**	**0.5976**	**495**	**81.68**	**0.6432**

Table 4
Misclassification (MC), Percentage of overall class Accuracy (PA) and *Kappa Index of Agreement (KIA)* values for BLOCKS data set.

Classification method	%training data								
	10			20			50		
	MC	PA	KIA	MC	PA	KIA	MC	PA	KIA
MLP	480	90.25	0.5732	386	91.18	0.6101	229	91.62	0.6428
Conventional NF	405	91.77	0.5981	334	92.36	0.6431	179	93.45	0.6671
Proposed NF	**265**	**94.61**	**0.6436**	**228**	**94.79**	**0.7001**	**127**	**95.35**	**0.7220**

Table 5
Misclassification (MC), Percentage of overall class Accuracy (PA) and *Kappa Index of Agreement (KIA)* values for SATIMAGE data set.

Classification method	%training data								
	10			20			50		
	MC	PA	KIA	MC	PA	KIA	MC	PA	KIA
MLP	1277	77.94	0.7456	1125	78.13	0.7553	655	79.63	0.7701
Conventional NF	1192	79.40	0.7578	1010	80.36	0.7694	602	81.28	0.7976
Proposed NF	**1036**	**82.10**	**0.7976**	**901**	**82.48**	**0.8103**	**529**	**83.55**	**0.8334**

3.2.2. PHONEME data

The aim of this data set is to distinguish between nasal and oral vowels (two classes) (ELENA Database, 1995). It contains vowels coming from 1809 isolated syllables (for example: pa, ta, pan,. . .). Five different attributes were chosen to characterize each vowel. They are the amplitudes of the five first harmonics. The data set has 5404 samples.

The performance comparison results are shown in Table 3. From the table it is seen that for 10% training data the proposed NF method provided nearly 5% increment of accuracy compared to the MLP. Correspondingly the MC values are 968 and 1245 using the proposed NF and MLP methods, respectively. The proposed NF method also maintained its supremacy over the conventional NF. These improvements are retained in case of 20% and 50% training data also. Using an analysis with KIA which provides the overall class-wise agreement, we found a superiority of the proposed method over MLP and conventional NF methods. At 10% training data, the KIA with the proposed NF, conventional NF and MLP methods are 0.5861, 0.4672 and 0.4012, respectively. The improvement of the proposed method is also obtained for both 20% and 50% training data.

3.2.3. BLOCKS data

The problem involved in this data set (listed as *Page Blocks Classification* (Asuncion & Newman, 2007)) is to classify the blocks of a page layout of a document that has been detected by a segmentation process. This is an essential step in document analysis in order to separate text from graphic areas. The five classes are: text (1), horizontal line (2), picture (3), vertical line (4) and graphic (5). BLOCKS data set has 10 features and 5 classes with 5473 sample patterns.

It is observed that the proposed NF classifier provides a better classification accuracy compared with those obtained with conventional NF and MLP. With 10% training data the MC and PA values for the proposed NF method are 265 and 94.61, respectively. These results are superior compared with the MC and PA values as 405 and 91.77, respectively obtained by conventional NF. In comparison with MLP the proposed NF method also provided improved results. It is seen that there is an increase of around 3.1% and 4.5% of accuracy and decrease of 145 and 215 in classification error with the use of proposed NF method compared with conventional NF and MLP. A similar trend of increment in the PA value is observed for both 20% and 50% training data by the proposed NF method. These results are depicted in Table 4. The table also reveals that with a minimum number of training data, the proposed method provided promising results compared with others with more training data.

In addition to the above comparison, we have made a class-wise overall accuracy measurement in terms of KIA and found that the KIA is clearly supporting the superiority of the proposed method. The KIA for the proposed NF, conventional NF and MLP are 0.6436, 0.5981 and 0.5732, respectively, for 10% training data. Similar improved KIA values are obtained using the proposed NF method for both 20% and 50% training data, which justified its better class-wise agreement compared with the other two methods.

3.2.4. SATIMAGE data

The SATIMAGE data set was generated from Landsat Multi-Spectral Scanner image data (listed as ***Statlog (Landsat Satellite)*** (Asuncion & Newman, 2007)). The data patterns used for the present investigation are a sub-area of a scene of 82×100 pixels. Each pixel value contains information from four spectral bands. The aim is to predict six different land cover classes present in the data set. The data set contains 6435 patterns with 36 attributes (4 spectral bands $\times$ 9 pixels in neighborhood). In our experiment we have used four features (17–20) only as recommended by the database designer (i.e., the four spectral values) (Asuncion & Newman, 2007).

The performance comparison of results in terms of MC and PA with this data set is shown in Table 5. It is observed that with 10%, 20% and 50% training data, the performance of the proposed NF classification method is better compared with conventional NF and MLP. For example, with 10% training data, the MC values are 1036, 1192 and 1277 using the proposed NF, conventional NF and MLP based classification methods, respectively. Similarly, the corresponding PA values for these classifiers are 82.10, 79.40

and 77.94. We have also measured the class-wise overall accuracy with the KIA index and observed that these values supported the improvement of the proposed method compared with the other two as mentioned above. Table 5 depicts the results in terms of KIA.

3.3. Classification of remote sensing images (i.e., partially labeled data sets)

In the present study, two multispectral remote sensing images (size 512×512) obtained from two different satellites (IRS-1A and SPOT) with different spatial and spectral resolution are used. The satellite images considered here contain a small number of labeled patterns. The actual classes (land covers) present in the original images (not shown in the manuscript) are not visible clearly. So we have displayed the enhanced images in Figs. 3(a) and 5(a), which highlight the different land cover regions properly. However the algorithms are implemented on the original images (not shown in the manuscript).

IRS-1A images

The IRS-1A image (shown in Fig. 3(a)) was obtained from Indian Remote Sensing Satellite (IRS Data Users Hand Book, 1989). We have used the images taken from the Linear Imaging Self Scanner (LISS-II). LISS-II has a spatial resolution of 36.25 m × 36.25 m and works in the wavelength range of 0.45–0.86 μm. The whole spectrum range is decomposed into four spectral bands, namely, blue band (band1), green band (band2), red band (band3) and near infrared band (band4) with wavelengths 0.45–0.52 μm, 0.52–0.59 μm, 0.62–0.68 μm and 0.77–0.86 μm, respectively.

The image in Fig. 3(a) covers an area around the city of Calcutta in the near infrared band having six major land cover classes. These are *pond or fishery water* (PW), *turbid water* (TW), *concrete area* (CA), *habitation* (HAB), *vegetation* (VEG) and *open spaces* (OS). The PW class contains pond water, fisheries etc. Sea water, river water etc., where the soil content is more, belong to TW class. CA class consists of buildings, roads, airport runways, bridges etc. Suburban and rural habitation, i.e., concrete structures are comparatively lower in density than the previous class (CA) and come under HAB class. VEG class essentially represents crop and forest areas. OS class contains barren land. More specifically, a pixel with less greenery and fewer concrete structures falls into this class.

SPOT image

The SPOT image shown in Fig. 5(a) is obtained from SPOT satellite (Systeme Pour d'Observation de la Terre) (Richards & Jia, 2006), which carries an imaging device HRV (High Resolution Visible). The Calcutta image used here has been acquired from the HRV that uses the wavelength range 0.50–0.89 μm. The whole spectrum range is decomposed into three spectral bands, namely, green band (band1), red band (band2) and near infrared band (band3) of wavelengths 0.50–0.59 μm, 0.61–0.68 μm and 0.79–0.89 μm, respectively. This image has a higher spatial resolution of 20 m × 20 m. We have considered the same six different classes for the land cover classification of the SPOT image. These are *pond or fishery water* (PW), *turbid water* (TW), *concrete area* (CA), *habitation* (HAB), *vegetation* (VEG) and *open spaces* (OS) as mentioned above.

3.3.1. Classification of IRS-1A Calcutta image

The classified (IRS-1A Calcutta) images obtained are shown in Fig. 3(b)–(d). From the visualization point of view, it is clear from the figures that the proposed NF classifier performed better in classifying the land cover (i.e., segregating different areas) compared to the conventional NF and MLP methods. Various regions or land cover classes in the IRS-1A Calcutta image are clearly identified from the classified image. From Fig. 3(d) (classified image using the proposed NF method) we see that the *Hooghly (Ganges)* river, situated in the middle of the image, belongs to TW class. The *pond or fishery water* (PW class) is easily identified from the classified image. The other classes such as CA, HAB, VEG and OS are also clearly visible. The above mentioned objects are more or less visible in case of the classified images obtained with conventional NF and MLP classifiers. A zoomed version of some classified regions such as *Saltlake area* (CA and HAB classes) and *pure water* class are shown in Fig. 4 to get an improved visualization.

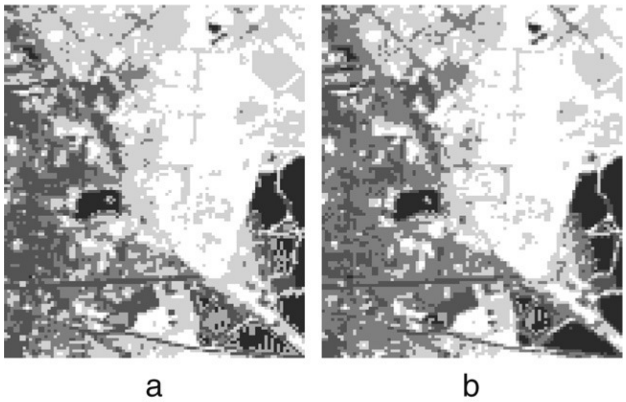

Fig. 4. Zoomed version of a selected region (*Saltlake area* and *pure water*) of classified IRS-1A Calcutta image with (a) MLP, and (b) proposed NF method.

It is observed that the *Saltlake area* is more clear, distinct and well shaped using the proposed NF method compared with the other two. Similarly, the separation of the *pure water* regions with the proposed method is comparatively more distinct. With the use of the proposed method, the classes became more separated and well identified. A concrete distinction between various classes obtained by different classifiers is properly justified with the estimation of quantitative indexes rather than only visualizing the regions.

We have used two performance measurement indexes (β and *DB*) to justify the findings obtained by visualization. As discussed in the previous section, for a fixed number of classes, the greater the homogeneity behavior within the class, the greater is the β value and Table 6 depicts the results of β to support this. As expected, the β value is the highest for the training data, i.e., 9.4212 for IRS-1A Calcutta image. Its values are 7.1587, 7.7535 and 8.6129 for the three classifiers, i.e., with MLP, conventional NF and proposed NF, respectively. From these values it is clear that the proposed NF classification system yields better results (*highest β value*) compared with the other two methods. As a whole we can establish the following β relation i.e.,

$$\beta_{training} > \beta_{proposed\ NF} > \beta_{conventional\ NF} > \beta_{MLP}.$$

From this relation we can get a gradation of performance quality of these three classification methods. We found that there is a gain (Eq. (8)) of 11.08% in β with the proposed NF method over conventional NF and 20.31% over MLP, which is highly appreciable. Similarly, the *DB* value is also supporting the superiority of the proposed classification method over the other two methods. The *DB* values are 0.7019, 0.8113 and 0.9390 for the proposed NF, conventional NF and MLP classification methods, respectively. As expected, the possible lowest *DB* value for the training data is found to be 0.5621. These results are depicted in Table 7. The corresponding gain obtained with respect to *DB* using the proposed NF over MLP and conventional NF are also evaluated. It is found that about 25% and 13% gains are achieved with the proposed NF method over the other two and the results are quite significant. Thus, it is clear that the proposed classification method is more suitable for the classification of the IRS-1A Calcutta image as it provides better compact and separable classified regions.

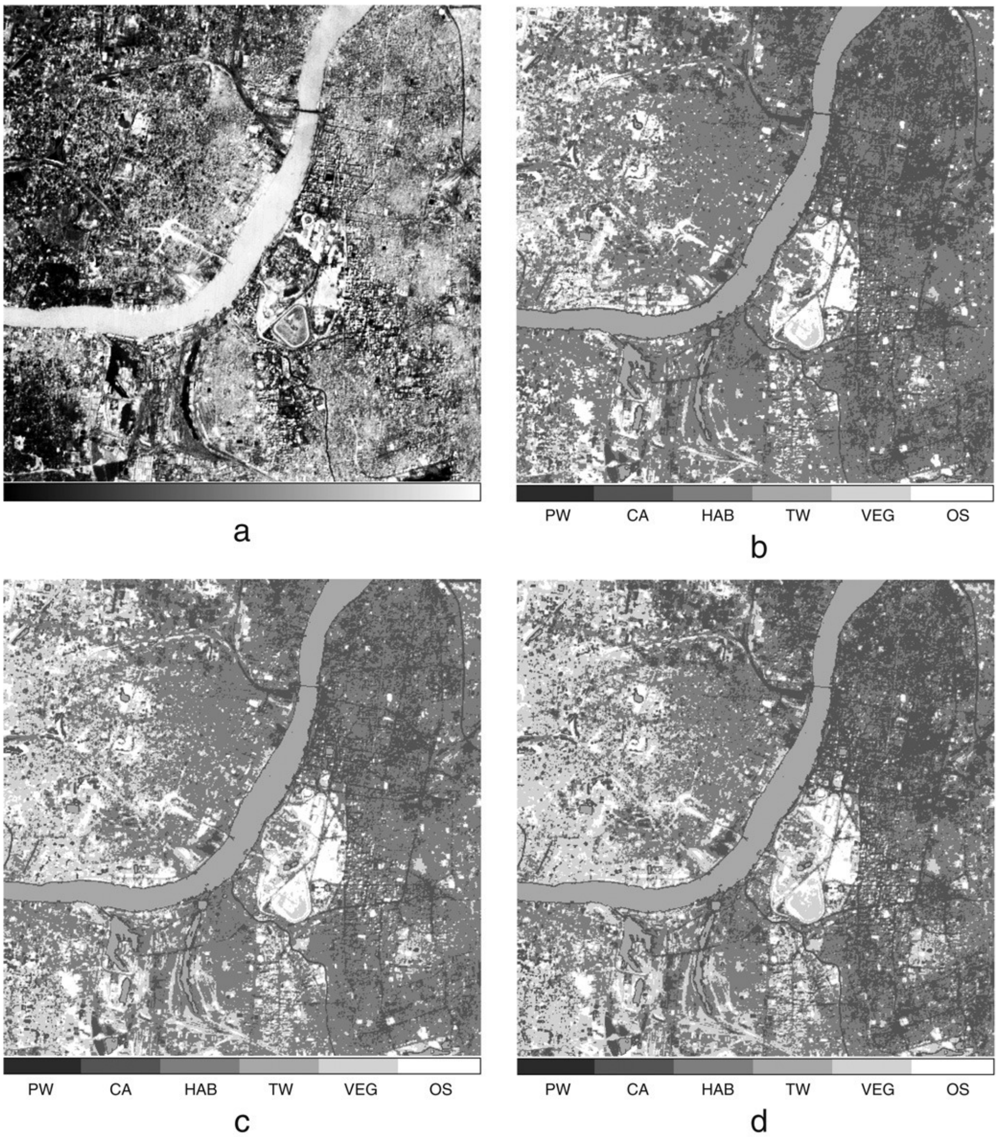

Fig. 5. SPOT Calcutta images: (a) Enhanced input, (b) land covers classified by MLP, (c) land covers classified by conventional NF, and (d) land covers classified by proposed NF methods.

Table 6
β value and corresponding gain for different classification methods.

Classification method	IRS Calcutta image	SPOT Calcutta image
Training patterns	9.4212	9.3343
MLP	7.1587	7.0542
Conventional NF	7.7535	7.6978
Proposed NF	**8.6129**	**8.5575**
Gain of proposed NF over MLP	**20.31%**	**21.31%**
Gain of proposed NF over conventional	**11.08%**	**11.16%**

Table 7
DB (Davies–Bouldin) value and corresponding gain for different classification methods.

Classification method	IRS Calcutta image	SPOT Calcutta image
Training patterns	0.5621	1.4943
MLP	0.9390	3.3512
Conventional NF	0.8113	2.4561
Proposed NF	**0.7019**	**1.9801**
Gain of proposed NF over MLP	**25.25%**	**26.71%**
Gain of proposed NF over conventional	**13.48%**	**19.38%**

3.3.2. *Classification of SPOT Calcutta image*

In case of SPOT Calcutta image, the classified images are shown in Fig. 5(b)–(d) for MLP, conventional NF and proposed NF classifiers, respectively. From the figures it is observed that there is a clear separation of different regions. The classified image shown in Fig. 5(d), using the proposed NF classifier, different classes or regions are more clearly identified compared with other two images (Fig. 5(b) and (c)), which are generated by other two methods. A zoomed version of some portions, e.g., two roads just above the bottom left of the image are shown in Fig. 6 to see the difference in the classified regions more clearly. From the figure, it is evident that the proposed method produced well structured and properly shaped regions (roads in CON class) compared with other methods considered in this investigation. However, a better performance comparison with the help of β and *DB* indexes can be seen from Tables 6 and 7. Table 6, shows that the β value for the training data set is 9.3343. Its values are 7.0542, 7.6978 and 8.5575 for the classified images using MLP, conventional NF and proposed NF classifiers, respectively. In this case an improved performance of the proposed NF classifier over others is also observed. It is seen

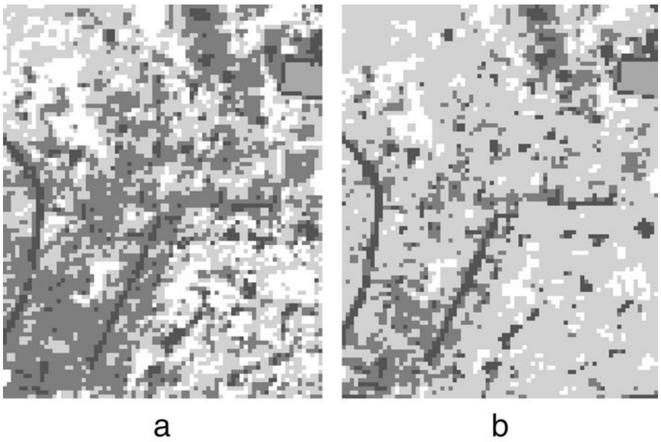

Fig. 6. Zoomed version of a selected region of classified SPOT Calcutta image with (a) MLP, and (b) proposed NF method.

that β is the highest for the proposed NF classification method that conveyed the message of its superiority. The same β *relation* in the classification of SPOT Calcutta image is observed as in the case of IRS-1A Calcutta. In this case, the gain obtained with the proposed NF classification method is 21.31% over MLP and 11.16% over conventional NF, which clearly justified the better classification ability of the proposed method. Similarly, the *DB* value as shown in Table 7, is the minimum in case of the proposed NF classification method compared with the other two methods. These values are 1.9801, 2.4561 and 3.3512 for the proposed NF, conventional NF and MLP based classification methods, respectively. The *DB* value for the training data is the lowest for the SPOT image and found to be 1.4943. The gain calculation on the basis of *DB* also supported the superiority of the proposed NF method and thus, develops more compact and separable land cover classes compared with other two methods.

From the classification results of four different conventional fully labeled data sets and two partially labeled multi-spectral remote sensing images, it is observed that for all cases the proposed NF method outperformed the conventional NF and MLP. The investigation revealed that with a lower percentage of training data also the proposed method performed well compared with the other two methods. Similarly, from the land cover classification of two partially labeled remote sensing images, we found that the proposed NF based classification method is superior as observed visually from the different classified regions. Performance comparison among these classifiers has been made using two quantitative indexes. The values of quantitative indexes also supported the superiority of the proposed NF classification method. It is justified that the classified regions using the proposed NF classification method are very distinct and the structures are more crisp, homogeneous, compact and easily separable compared with the conventional NF and MLP methods. However, the computational complexities of the proposed NF classification system will increase in data sets having a greater number of features and classes. Although after fuzzification we will have more features, since the system learns from a lower number of training patterns, the total time required will not be very large.

4. Conclusion

We have proposed a novel neuro-fuzzy model for classification and demonstrated successfully its effectiveness for classification of fully and partially labeled patterns. The method exploits and incorporates the basic advantages of neural networks such as massive parallelism, robustness, adaptivity and optimality in one hand; and impreciseness and uncertainty handling capability of fuzzy sets on the other hand. Besides these generic advantages, the proposed model develops a *membership matrix* that provides information of feature-wise degree of belonging of a pattern to all classes instead to a particular class. This in turn provides better generalization capability.

Various performance measures such as the number of misclassifications, percentage of classification accuracy, Kappa index of agreement for completely labeled data sets and quantitative indexes such as the β index of homogeneity and the *DB* index of compactness for partially labeled data sets are used to justify the promising performance of the proposed method. The percentage gain obtained in β and *DB* values corroborates these finding. It is observed that the proposed classification method provides an improved performance even with less training data. In case of remote sensing images, the performance is very high (*gain is more than* 20%).

The computational complexity of the proposed NF classification model is little high. However, its learning ability with small percentage of training samples will make it practicably applicable to problems with a large number of classes and features. In future, we plan to compare this NF technique with some other existing NF techniques, in all aspects such as time complexity and classification accuracy.

Acknowledgements

The authors would like to thank the reviewers for their valuable and constructive suggestions. Thanks are also due to the Department of Science and Technology, Government of India, 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.

References

Abe, S. (2001). *Pattern classification: Neuro-fuzzy methods and their comparison*. Springer Verlag.

Acharyya, M., De, R. K., & Kundu, M. K. (2003). Segmentation of remotely sensed images using wavelet features and their evaluation in soft computing framework. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(12), 2900–2905.

Anthony, M., & Bartlett, P. L. (1999). *Neural network learning: Theoretical foundations*. Cambridge University Press.

Asuncion, A., & Newman, D. (2007). UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences. http://www.ics.uci.edu/~mlearn/MLRepository.html.

Baldi, P. F., & Homik, K. (1995). Learning in linear neural networks: A survey. *IEEE Transactions on Neural Networks*, *6*(4), 837–858.

Baraldi, A., Binaghi, E., Blonda, P., Brivio, P. A., & Rampini, A. (2001). Comparison of the multilayer perceptron with neuro-fuzzy techniques in the estimation of cover class mixture in remotely sensed data. *IEEE Transaction on Geoscience and Remote Sensing*, *39*(5), 994–1005.

Boskovitz, V., & Guterman, H. (2002). An adaptive neuro-fuzzy system for automatic image segmentation and edge detection. *IEEE Transaction on Fuzzy Systems*, *10*(2), 247–262.

Card, D. H. (1982). Using known map category marginal frequencies to improve estimates of thematic map accuracy. *Photogrammetric Engineering and Remote Sensing*, *48*(3), 431–439.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*(1), 35–46.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *1*(4), 224–227.

ELENA Database (1995). *Deliverable R3-B1-P Task B1: Databases* (ESPRIT Basic Research Project Number 6891), MLG, UCL, Belgium. http://www.dice.ucl.ac.be/mlg/DataBases/ELENA/REAL/.

Gamba, P., & Dellacqua, F. (2003). Increased accuracy multiband urban classification using a neuro-fuzzy classifier. *International Journal of Remote Sensing*, *24*(4), 827–834.

Ghosh, A., Meher, S. K., & Shankar, B. U. (2008). A novel fuzzy classifier based on product aggregation operator. *Pattern Recognition*, *41*(3), 961–971.

Ghosh, A., Pal, N. R., & Pal, S. K. (1993). Self-organization for object extraction using a multilayer neural network and fuzziness measures. *IEEE Transaction on Fuzzy Systems*, *1*(1), 54–69.

Han, J., Lee, S., Chi, K., & Ryu, K. (2002) Comparison of neuro-fuzzy, neural network, and maximum likelihood classifiers for land cover classification using IKONOS multispectral data. In: *Proceedings of the IEEE international symposium on geoscience and remote sensing*, pp. 24–28.

Haykin, S. (1997). *Neural networks: A comprehensive foundation* (2nd ed.). Prentice Hall.

IRS Data Users Hand Book. (1989). Document no. IRS/NRSA/NDC/HB-02/89, NRSA, Hyderabad, India.

Keller, J. K., & Hunt, D. J. (1985). Incorporating fuzzy membership functions into the perceptron algorithm. *IEEE Transaction on Pattern Analysis Machine Intelligence*, *7*(6), 693–699.

Kuncheva, L. I. (2000). *Fuzzy classifier design*. Springer-Verlag.

Kwon, K., Ishibuchi, H., & Tanaka, H. (1994). Neural networks with interval weights for nonlinear mapping of interval vectors. *IEICE Transactions on Information and Systems*, *E77-D*, 409–417.

Marinai, S., Gori, M., & Soda, G. (2005). Artificial neural networks for document analysis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(1), 23–35.

Mitra, P., Shankar, B. U., & Pal, S. K. (2004). Segmentation of multispectral remote sensing images using active support vector machines. *Pattern Recognition Letters*, *25*(12), 1067–1074.

Pal, S. K., & Ghosh, A. (1996). Neuro-fuzzy computing for image processing and pattern recognition. *International Journal of Systems Science*, *27*, 1179–1193.

Pal, S. K., Ghosh, A., & Kundu, M. K. (Eds.) (2000). *Soft computing for image processing*. Heidelberg: Physica-Verlag.

Pal, S. K., Ghosh, A., & Shankar, B. U. (2000). Segmentation of remotely sensed images with fuzzy thresholding, and quatitative evaluation. *International Journal of Remote Sensing*, *21*(11), 2269–2300.

Pal, S. K., & Majumder, D. D. (1977). Fuzzy sets and decision making approaches in vowel and speaker recognition. *IEEE Transaction on Systems, Man and Cybernetics*, *7*(1), 625–629.

Pal, S. K., & Majumder, D. K. D. (1986). *Fuzzy mathematical approach to pattern recognition*. Wiley Eastern Ltd.

Pal, S. K., & Mitra, S. (1999). *Neuro-fuzzy pattern recognition: Methods in soft computing*. New York, NY, USA: John Wiley & Sons, Inc.

Pedrycz, W. (1990). Fuzzy sets in pattern recognition: Methodology and methods. *Pattern Recognition*, *23*(1/2), 121–146.

Qiu, F., & Jensen, J. R. (2004). Opening the black box of neural networks for remote sensing image classification. *International Journal of Remote Sensing*, *25*(9), 1749–1768.

Richards, J. A., & Jia, X. (2006). *Remote sensing digital image analysis: An introduction* (4th ed.). Berlin, Heidelberg: Springer-Verlag.

Ripley, B. (1996). *Pattern recognition and neural networks*. Cambridge University Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Tso, B., & Mather, P. M. (2001). *Classification methods for remotely sensed data*. London: Taylor and Francis.

Varshney, P. K., & Arora, M. (2004). *Advanced image processing techniques for remote sensed hyperspectral data*. Germany: Springer-Verlag.

Zadeh, L. A. (1965). Fuzzy sets. *Information Control*, *8*(3), 338–353.

Zhang, G. P. (2000). Neural networks for classification: A survey. *IEEE Transactions on Systems, Man, and Cybernetics (Part C: Applications and Reviews)*, *30*(4), 451–462.

Zurada, J. M. (1992). *Introduction to artificial neural systems*. PWS Publishing Company.