International Journal of Wavelets, Multiresolution
and Information Processing
Vol. 5, No. 4 (2007) 589–611

NEURO-WAVELET CLASSIFIER FOR MULTISPECTRAL REMOTE SENSING IMAGES

B. UMA SHANKAR*, SAROJ K. MEHER† and ASHISH GHOSH‡

Machine Intelligence Unit, Indian Statistical Institute
203 B. T. Road, Kolkata 700108, India
**uma@isical.ac.in*
†*saroj_t@isical.ac.in*
‡*ash@isical.ac.in*

Received 3 June 2006
Revised 23 December 2006

A neuro-wavelet supervised classifier is proposed for land cover classification of multispectral remote sensing images. Features extracted from the original pixels information using wavelet transform (WT) are fed as input to a feed forward multi-layer neural network (MLP). The WT basically provides the spatial and spectral features of a pixel along with its neighbors and these features are used for improved classification. For testing the performance of the proposed method, we have used two IRS-1A satellite images and one SPOT satellite image. Results are compared with those of the original spectral feature based classifiers and found to be consistently better. Simulation study revealed that Biorthogonal 3.3 (Bior3.3) wavelet in combination with MLP performed better compared to all other wavelets. Results are evaluated visually and quantitatively with two measurements, β index of homogeneity and Davies–Bouldin (DB) index for compactness and separability of classes. We suggested a modified β index in accessing the percentage of accuracy (PA_β) of the classified images also.

Keywords: Remote sensing; land cover classification; wavelet transform; neural networks.

Mathematics Subject Classification 2000: 22E46, 53C35, 57S20

1. Introduction

Classification of different land cover regions of remote sensing images is essential for efficient interpretation of them. This task is very complex. The reason behind this is the poor illumination quality and low spatial resolution of remotely placed sensors and rapid changes in environmental conditions. Basically, the regions such as vegetation, soil, water-bodies, concrete structures etc. of a natural scene are often not well separated, leading to overlapping regions. Moreover, the gray value

assigned to a pixel is an average reflectance of different types of land cover classes present in the corresponding pixel area.

Multispectral remotely sensed images comprise information over a large range of variation on frequencies (information), and these frequencies change over different regions (i.e. non-stationary behavior of the signal), which needs to be estimated properly for improved classification. The multispectral remote sensing image data have both spectral features with correlated bands and spatial features correlated in the same band (also known as spatial correlation). An efficient utilization of these spectral and spatial (contextual) information can improve the classification performance significantly compared to the conventional non-contextual information based classical (statistical) and modern (fuzzy, neural, and kernel based techniques including support vector machines) methods. Most of the conventional multispectral remote sensing image classification systems detect object classes only using the spectral information of the individual pixel/pattern, while a large amount of spatial information of neighboring pixels are neglected. Hence, the pixels are classified based on its spectral intensities and does not give attention to its spatial dependencies. Important information from neighboring pixels is thus ignored. Such approaches may be reasonable if spatial resolution is high or when the spectral intensities are well separated for different classes, which is rarely found in any real life data. For example, in the urban areas, the densities of the spectral intensities are seldom well separated, e.g. various classes like concrete structure, vegetation, habitation, roads, small and narrow water bodies have a lot of overlapping regions with each other. Thus, it is important to decide whether the spatial arrangements of data or a transformation of it to a different space where the spatial information is uncorrelated can be used as features.

Much research effort have been made to take the advantages of neighboring pixel information[1−6] and applied for the classification of remotely sensed data. These include texture features extracted from angular second moments, contrast, correlation, entropy, variance, etc., computed from the gray level co-occurrence matrices.[2] The extracted textural features play an important role and it may increase the performance of the classifier. However, these methods are computationally expensive due to the estimation of autocorrelation parameters and transition probabilities. Also, the texture elements are not easy to quantify and it deals with the spatial distribution of the gray levels over a portion of an image. Later on, Gaussian Markov random fields (GMRF)[7−9] and Gibbs random fields[10] were proposed to characterize textures. Further, local linear transformations are also used to compute textural features.[11] The above-mentioned conventional statistical approaches to texture analysis are restricted to the analysis of spatial interactions over relatively small neighborhoods on a single scale.

Normally the multispectral pixel information are usually inefficient in discriminating properly the different land cover classes having ill-defined and overlapping classes. Hence, use of both spectral and textural information in classification problems is more appropriate. Attempts were made to use both spectral[6,12] and textural features[2−5] independently. The conventional textural feature extraction

methods normally assume the stationarity of the texture within the analysis range and hence may not be useful. For this reason, the characterization of textures should be in both spatial and frequency domains, simultaneously.

One efficient way to deal with such problems is to recognize the image by a number of subsampled approximations of it at different resolutions and desired scales. This approach is conventionally called multiresolution analysis.[13,14] In this regard Gabor filters and wavelet transform (WT) received a lot of attention.[15–18] However, a major disadvantage in using Gabor transform is that the output of Gabor filter banks are not mutually orthogonal, which may result in extracting features having significant correlation between them. Moreover, these transformations are usually not reversible, which limits their applicability for textural feature extraction. By using WT, most of these disadvantages can be avoided thereby providing a precise and unifying framework for the analysis and characterization of a signal at different scales.[15] Another advantage of WT over Gabor filters is that the low pass and high pass filters used in the WT remain the same between two consecutive scales while the Gabor approach requires filters of different parameters.[16] In other words, Gabor filters require proper tuning of filter parameters at different scales.

As discussed above, in order to capture the local variations in the orientation and frequency of texture elements that lead to the large scale frequency variation behavior of remotely sensed image textures, we need a joint spatial/spatial-frequency representation. In this regard, WT is found to be a promising tool for texture analysis in both spatial and spatial-frequency domains[19] simultaneously, as it has the ability to examine the signal at different scales. This means that the WT provides information in both spatial and spatial-frequency domains. The WT scheme thus analyzes the coarse image first and gradually increases the resolution to analyze the finer details. Basically the WT coefficients represent the characteristic in frequency bands indicating the characteristic like frequency and spatial location of the original pixel, from where the WT coefficients are originated. Thus, the WT coefficients extract neighboring pixels information that are uncorrelated in spatial domain.

Hence, the use of classification methods with these coefficients (pixels with extracted features) instead of original pixel value is more justifiable. These characteristics of the WT motivated us to use it for extraction of hidden features from remote sensing images with non-stationary behavior of the pixels. Research works related to texture classification using WT has already been carried out.[15,17] Use of statistical correlation based features in the WT domain for classification have also been reported.[18,20] In another study, WT has been used for target classification of remote sensing images.[21–25] A comparative study of multiwavelet,[26,27] wavelet, and shape features[2] for microcalcification classification in mammogram has been described in Ref. 28, where it is experimentally shown that the results of multiwavelet based approach were better compared to wavelet and statistical based features. However, the computational complexities of multiwavelet based classification is much higher than the wavelet based one.

Neural networks (NNs) are aimed to emulate 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.[29–32] Interactions among the neurons is very high in NNs making them suitable for making collective decisions. The main characteristics of NNs, namely, adaptivity, fault tolerant, robustness and optimality play important roles particularly in the field of pattern classification. The pattern classification problem becomes more complex when there is no prior information on class distribution. In such cases, NN-based systems use adaptive learning procedures that learn from examples and attempt to find a useful relation between input and output. Similarly, NNs are also useful to model complex nonlinear boundaries and to discover important underlaying regularities in the problem domain. In a conventional neural classification model, number of nodes in the input-layer is equal to the number of features present in the data pattern and the number of output-layer nodes is equal to the number of classes. Thus, the importance of different features will automatically be encoded in the connecting links during training and the nonlinear decision boundaries are generated to label the assigned classes by making collective decision. Preliminary attempts to use neural networks for land cover classification of remote sensing imagery are available.[33–37]

In the present work, we have tried to explore the advantages of WT instead of multiwavelet and statistical approaches by incorporating it as a preprocessor in the classification process. We have considered a conventional feed-forward neural network (MLP)[29,38] based classifier in the proposed scheme for land cover classification of remote sensing images. In this method, we first extract features of the input patterns/pixels using WT and use these features for the next step of classification. We have evaluated the performance of the proposed wavelet feature (WF) based neural classification method with different wavelets. The proposed classifier is tested on three multispectral remote sensing images: two four-band Indian Remote Sensing 1A (IRS-1A) satellite images and one three-band SPOT satellite image[39] for land cover classification. Comparison of results showed that the proposed WF based neural classification scheme yields superior results compared to spectral feature based NN classifier. The performance of the proposed neuro-wavelet (NW) classifier is further increased with bi-orthogonal wavelets.

Organization of the article is as follows. Section 2 describes the wavelet transform based feature extraction method. A brief description of neural classifier is made in Sec. 3. Section 4 discusses the performance evaluation measures. Comparative results with discussion are included in Sec. 5. Finally concluding remarks and scope for future work are given in Sec. 6.

2. Wavelet Transformation Based Feature Extraction

To deal with the non-stationary behavior of signals in an appropriate way, much research effort have been made so that the disadvantages of the Fourier transform

(FT),[14,19] which assumes the signal to be stationary within its total range of analysis, can be compensated. This problem can be sorted out to some extend using short time Fourier transform (STFT). However, the selection of window is a problem in STFT, as narrow windows give good time resolution but poor frequency resolution and wide windows give good frequency resolution but poor time resolution. Furthermore, wide windows may violate the condition of stationarity. The WT on the other hand is an extension of these transforms,[13,14,19,40] which does not lose the spatial localization of the signal frequencies and it can be fitted to signals by translation and dilation of the wavelet basis function (mother wavelet). The basic difference of the WT from FT is that the time localization of the signal frequencies will not be lost.

WT extends single scale analysis to multiscale analysis. The multiscale behavior of the WT analyzes or decomposes the signal in multiple number of scales, where each scale represents a particular coarseness of the signal, thus, the decomposition steps divide the signal into a set of signals of varying coarseness ranging from low frequency to high frequency components. Accordingly, WT tries to identify both the space and frequency information of the events simultaneously. This makes WT useful for signal feature analysis. This property is more useful for remote sensing data analysis where, any characteristic of the scene is first analyzed using low resolution and then an area of interest is analyzed in detail using an appropriate higher level resolution.

Since we are dealing with irregular textures of remotely sensed images, the decomposition of the signal into different scales, which can uncorrelate the data, as much as possible without losing their distinguishable characteristics, is particularly more useful when the WT is done with an orthogonal basis.[13] The distinguishable characteristics of the original information preserved in the WT decomposition are spatio-geometrical information, normally called the signature of the land cover. The orthogonal basis is more compact in its representation, as it allows the decomposition of the underlying space into orthogonal subspaces, which makes it possible to ignore some of the decomposed signals. This type of WT is also suitable for operations like feature extraction,[41,42] parameter estimation[43,44] and exact reconstruction of the signal series because of its invertible properties.[13,14]

WT is identical to a hierarchical subband system, where the subbands are logarithmically spaced in frequency. Wavelets are functions used in the transformation that act as a basis for representing many functions of the same family. A series of functions can be generated by translation and dilation of these functions called mother wavelets $\psi(x)$. The translation and dilation of the mother wavelet can be done by

$$\psi_{\gamma,\tau}(t) = |\gamma|^{-1/2}\psi\left(\frac{t-\tau}{\gamma}\right), \quad \gamma \neq 0 \text{ and, } \gamma \varepsilon \mathbb{R}, \ \tau \varepsilon \mathbb{R}, \tag{2.1}$$

where τ and γ are the translation and dilation parameters.

We have performed classification of remote sensing images using wavelets from different (Daubechies, Biorthogonal, Coiflets, Symlets) groups.[40] However, results are given for three wavelets as their performances are comparatively (empirically) better than others. These are Daubechies 3 (Db3), Biorthogonal 3.3 (Bior3.3) and Biorthogonal 3.5 (Bior3.5) wavelets.[40] In Daubechies wavelets, N (3 here) is related to support width, filter length, regularity and number of vanishing moments for the wavelets. For example, DbN is having support width of $2N-1$, filter length (number of filter coefficients) of $2N$, regularity about $0.2N$ for large N and number of vanishing moments is N. Similarly, 3.3 and 3.5 of biorthogonal wavelets indicate some properties like regularities, vanishing moments, filter length, etc. These wavelets are implemented with the multiresolution scheme given by Mallat,[13] which is briefly described below.

2.1. *Discrete WT and multiresolution analysis*

2.1.1. *1-Dimensional*

The discrete WT analyzes the signal at different frequency bands with different resolutions by decomposing the signal into low frequency (approximation) and high frequency (details) band information. The decomposition of the signal into different frequency bands is obtained by successive highpass and lowpass filtering of the signal. The 1-dimensional (1D) original signal (e.g. $s_n = [s_{nk}]$) is decomposed using a highpass filter h_k and a lowpass filter g_k, where $n = 1, 2, \ldots, Q$ is the number of decomposition levels and $k = 1, 2, \ldots, S$ is the number of signal samples. The decomposed signals consist of two parts $\mathbf{s}_{n-1}$ and $\mathbf{d}_{n-1}$, called the smooth or lowpass, and fine or highpass information. The two decomposed signals can be expressed as

$$s_{n-1,j} = \sum_{k=1}^{S} s_{nk} \cdot g_{k-2j}, \tag{2.2}$$

$$d_{n-1,j} = \sum_{k=1}^{S} s_{nk} \cdot h_{k-2j}, \tag{2.3}$$

with j as the length of the convolution mask. The reconstruction of the original signal from the decomposed wavelet coefficients can be performed as

$$s_{nk} = \sum_{j} [s_{n-1,j} \cdot \tilde{g}_{k-2j} + d_{n-1,j} \cdot \tilde{h}_{k-2j}], \tag{2.4}$$

where $\tilde{g}_k$ and $\tilde{h}_k$ are the reconstruction filter coefficients. The notation $(\downarrow 2)\mathbf{y}$ in Fig. 1 denotes the *downsampled* version of the sequence $\mathbf{y}$ by 2, i.e. $((\downarrow 2)\mathbf{y})_k = y_{2k}$. Here the odd-numbered wavelet coefficients are dropped and the even-numbered

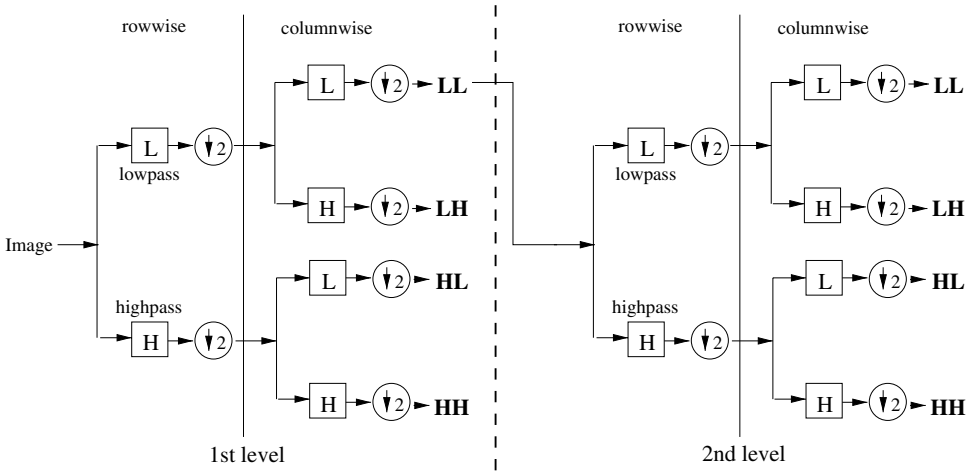


Fig. 1. Two-level discrete wavelet transform.

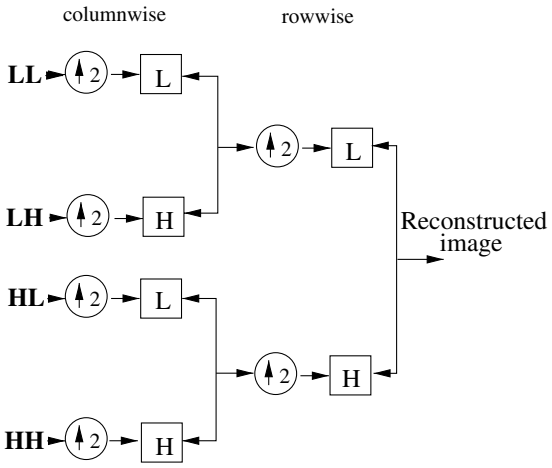


Fig. 2. Reconstruction of a single band sub-image using inverse WT.

wavelet coefficients are renumbered. The notation $(\uparrow 2)\mathbf{y}$ in Fig. 2 denotes the *upsampled* version of the sequence $\mathbf{y}$ by 2, i.e.

$$((\uparrow 2)\mathbf{y})_{2k} = y_k,$$

$$((\uparrow 2)\mathbf{y})_{2k+1} = 0.$$

Here, a zero is inserted between each pair of adjacent wavelet coefficients.

In the reverse process of discrete WT, normally called inverse WT, the reconstructed signal may not be exactly the same as the original one; however, it will be a good approximation. Due to these properties, WT is widely used for signal and more generally in image compression.[45,46]

The above description is for one step of discrete WT. The multiple levels of decomposition can be performed as

$$\begin{aligned} \mathbf{s}_n &\rightarrow \mathbf{s}_{n-1}, \mathbf{d}_{n-1} \\ \mathbf{s}_{n-1} &\rightarrow \mathbf{s}_{n-2}, \mathbf{d}_{n-2} \\ \ldots &\rightarrow \ldots . \end{aligned} \tag{2.5}$$

The decomposition operation in the WT halves the spatial resolution, since only half of the number of samples now characterize the entire signal. However, this operation doubles the frequency resolution, since the frequency band of the signal now spans only half the previous frequency band, effectively reducing the uncertainty in the frequency. The above procedure, which is also known as the subband coding, can be repeated for further decomposition. At every level, the filtering and *downsampling* will result in halving the number of samples (and hence halving the spatial resolution) and halving the frequency band span (and hence doubling the frequency resolution). This process of subband coding is also known as multiresolution analysis.[14,47]

2.1.2. *2-Dimensional*

The two-dimensional (2D) WT is performed by consecutively applying 1D WT on rows and columns of the 2D data. A 2D WT, which is a separable filter bank in row and column directions, decomposes an image into four sub-images.[13,14] Figure 1 shows this dyadic 2-level decomposition of a 2D image. H and L in Fig. 1 denote highpass and lowpass filters, respectively. $\downarrow 2$ denotes the *downsampling* (decrease the sample occurrence rate) by a factor of two. Thus in one level decomposition of the 2D WT, four sub-sampled versions of the original image are obtained. Among them, one contains the WT coefficients in low frequency range called the approximation part, i.e. LL (Fig. 1) and three in high frequency range in three directions: vertical (LH), horizontal (HL) and diagonal (HH), called the detail parts. The wavelet coefficients in these subbands provide frequency information of the original signal in four different frequency bands maintaining their spatial position the same as in the original. In addition to this, the coefficient values also preserve the neighborhood information. The decomposition of the wavelet coefficients can be extended to more than one level. The next level of decomposition can be performed on the approximation coefficient of the previous level. The other coefficients are assumed to be redundant as they contain information in the high frequency bands and normally provide less information. However, the next step of decomposition can be performed on all the coefficients obtained from the previous level. This decomposition method is called wavelet packet[14] decomposition.

In the higher level of decomposition, detailed information can be obtained. Thus the level of decomposition depends on the type of requirement and it varies with the problem in hand. To have an objective evaluation, we computed the average entropy,[48] which provides a measure of information, of the images for each level. We found that the average entropy value is not changing significantly after a certain level of decomposition; and we decided to decompose up to that level. For the present experiments, we stopped decomposition after only the second level, as the entropy measure was not changing much after this and thus we were not getting much extra information after this level of decomposition, even though the cost of computation kept increasing.

In the present experiment, we decomposed the approximation coefficients (LL image) of each level into four sub-images iteratively. We used orthogonal wavelet bases and thus there is hardly any loss or redundancy of information. From these wavelet coefficients, the corresponding reconstructed images are obtained using inverse WT, which will be used subsequently as the extracted features of the original image for classification purpose. Thus, we will have many extracted features from the input images and the number of images (features) will depend on the level of decomposition.

2.2. *Feature extraction*

As discussed above, the decomposed WT coefficients at different levels in different sub-images represent the information of the original pixel values as well as the information of the neighboring pixels. We have used these coefficients to construct features. In this regard, different bands of images are decomposed into the desired level (second level for the present experiments) using the 2D WT, which provides four subband images from each band. As a whole, 16 subband images can be obtained from a four-band image (original input) after one level of decomposition. It becomes 28 band sub-images with two levels of decomposition, and so on. The pixels of the sub-images are reconstructed to get the image information from the corresponding subband. The sub-images are then cascaded as shown in the Fig. 3 so that the extracted feature vectors of the original multispectral image can be obtained for the next step of classification. Cascading of different bands for generation of feature vector with Q-level of decomposition can be performed as

$$(I^1_{LL-Q}, I^1_{LH-Q}, I^1_{HL-Q}, I^1_{HH-Q}, \ldots, I^1_{LH-1}, I^1_{HL-1}, I^1_{HH-1}, \ldots,$$
$$I^B_{LL-Q}, I^B_{LH-Q}, I^B_{HL-Q}, I^B_{HH-Q}, \ldots, I^B_{LH-1}, I^B_{HL-1}, I^B_{HH-1}),$$

where, I^B_{LH-1} denotes the sub-image at first level for first band with B as the number of spectral bands of the original image. Hence the feature vector of each pattern of the above decomposition will be of length $B(3Q+1)$. Thus, a two-band multispectral image with three-levels of decomposition creates a feature vector of length 20. Figure 3 shows the cascading of sub-images of a single band image, which is the feature vector of a classifier (here a neural network).

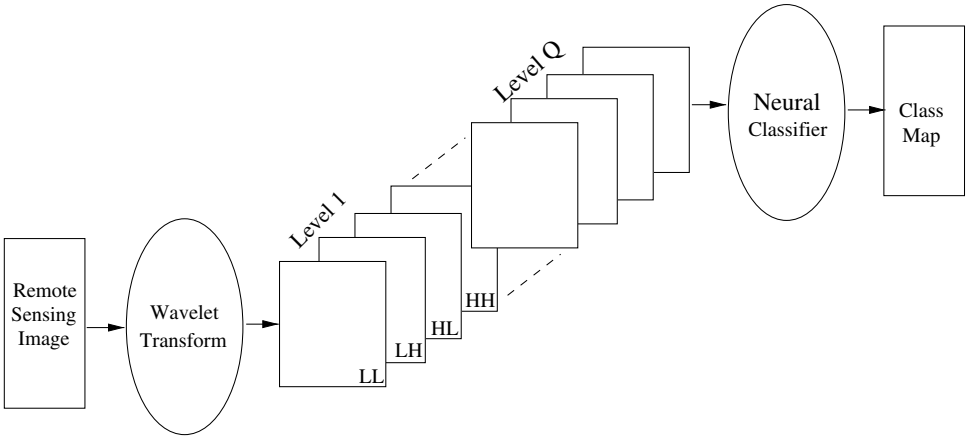


Fig. 3. Classification procedure.

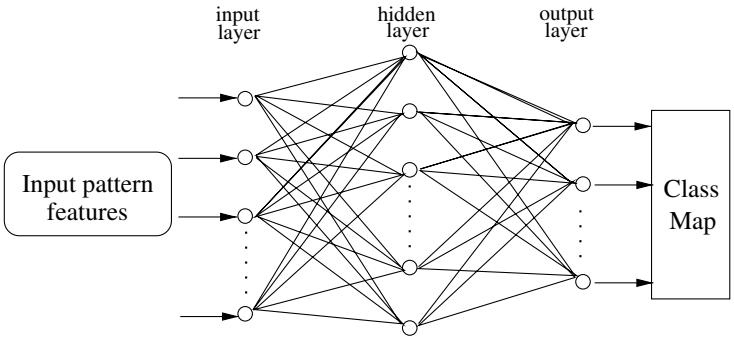


Fig. 4. A three-layer feed-forward neural network.

3. Neural Classifier

The WF (wavelet features) generated by WT are used as input to NNs. The proposed WF based classification method has been implemented using the most popular feed forward multi-layer perceptron (MLP, shown in Fig. 4) classifier. The MLP (used in the experiment) has three layers- known as input, hidden and output layers, respectively. Each processing node, except the input-layer nodes, calculates a weighted sum of the outputs from the nodes in the preceding layer to which it is connected. This weighted sum then passes through a transfer function to derive its own output, which is then fed to the nodes in the next layer. Thus the input to node v is obtained as

$$\text{net}_v = \sum_u W_{uv} O_u + \text{bias}_v \tag{3.1}$$

and output

$$O_v = S(\text{net}_v), \tag{3.2}$$

where w_{uv} is the weight for the connection linking node u to node v, bias_v is the bias value for node v, O_u is the output of node u, and S stands for the activation function (AF). Here the AF is considered as a sigmoid function[29] and is defined as

$$S(\text{net}_v) = \frac{1}{1 + e^{(-\text{net}_v)}}. \tag{3.3}$$

MLP uses back-propagation (BP) learning algorithm[38] for weight updating. The BP algorithm basically reduces the sum of squared error called cost function (CF) between the actual and desired output of output-layer neurons in a gradient descent manner. The CF is given as

$$CF = \frac{1}{2}\sum_g \sum_v (O_{gv} - t_{gv})^2, \tag{3.4}$$

where g is a training pattern and v is an output node. O_{gv} denotes the output of node v when the training pattern g is applied to the network, and t_{gv} is the corresponding target output. The error parameter value thus generated is propagated backward to correct the weights following Eq. (3.5).

$$\Delta W_{vu}(n+1) = \alpha \Delta W_{vu}(n) + \eta \delta_v O_u, \tag{3.5}$$

where n, α, η and δ are the iteration number, momentum parameter, learning rate and node error, respectively. The details of BP algorithm including derivation of the equation can be obtained from Ref. 29. The number of nodes in the input-layer is equal to the number of features and the number of nodes in the output-layer is equal to the number of classes present in the data set. The hidden-layer nodes, chosen in the present experiment, is equal to the square root of the product of the number of input and output layer nodes.[38] We have used a single hidden-layer for the present investigation.

4. Performance Measurement Indices

Selection of the training samples for all classes are made according to a prior assumption of the land cover regions. These training samples are used to estimate the parameters of the neural classifier. After training the classifier, it is used to classify the land covers of the whole image. In the following text, we discuss the performance measures used for the present investigation.

4.1. *β index*

When the labels of all the pixels in an image are known, the Kappa index[49] can be used to estimate the classification accuracy. But here the case is different, i.e. the class label of all the pixels are not known. For this case, the cluster validity

index[50] for measuring the segmentation quality is more appropriate. We used this quantitative index (QI), called β, to estimate the performance of the classifiers. This measure has been successfully used in the assessment of segmentation quality.[50–52] The β in Eq. (4.1) is defined as the ratio of the total variation and within-class variation of the data set.[50] Since the numerator is constant for a given image, the 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 given number of classes, *the higher the homogeneity within the extracted classes, the higher will be the* β *value*. Mathematically, β can be represented as

$$\beta = \frac{\sum_{i=1}^{C}\sum_{j=1}^{M_i}(\mathbf{x}_{ij} - \overline{\mathbf{x}})^2}{\sum_{i=1}^{C}\sum_{j=1}^{M_i}(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)^2}, \tag{4.1}$$

where $\overline{\mathbf{x}}$ 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 grey value of the jth pixel ($j = 1, 2, \ldots, M_i$) in class i, and $\overline{\mathbf{x}}_i$ is the mean of M_i pixel values of the ith class. The higher the β value the better is the partitioning.

4.2. *Davies–Bouldin index*

The Davies–Bouldin (DB) index of cluster validation has been defined and used in Ref. 53. Also, various researchers have used it for cluster validation in the past. However, 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-cluster separation as well as intra-cluster 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. Hence DB index is a function of three parameters, i.e. S_i, S_j and d_{ij}. The dispersion S_i of ith cluster and the separation d_{ij} between the ith and the jth clusters are defined as

$$S_{i,q} = \left(\frac{1}{|X_i|}\sum_{\mathbf{x}\varepsilon X_i} ||\mathbf{x} - \mathbf{v}_i||_2^q\right)^{\frac{1}{q}}, \tag{4.2}$$

and

$$d_{ij,t} = \left[\sum_{s=1}^{p} |v_{si} - v_{sj}|^t\right]^{\frac{1}{t}}. \tag{4.3}$$

$S_{i,q}$ 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. $S_{i,2}$ is the square root of the mean square error

of the points in the extracted ith class with respect to the centroid of it, and so on. $d_{ij,t}$ is the Minkowski distance of order t between the centroids that characterizes the extracted classes i and j. In the present experiment, we have taken $q = t = 2$. Hence we compute

$$R_{i,qt} = \max_{j,j\neq i}\left[\frac{S_{i,q} + S_{j,q}}{d_{ij,t}}\right]. \tag{4.4}$$

The DB index is then defined as

$$DB = \frac{1}{C}\sum_{i=1}^{C} R_{i,qt}, \tag{4.5}$$

with C as the number of clusters/classes. The smaller the DB value, the better is the partitioning.[53]

5. Results and Discussion

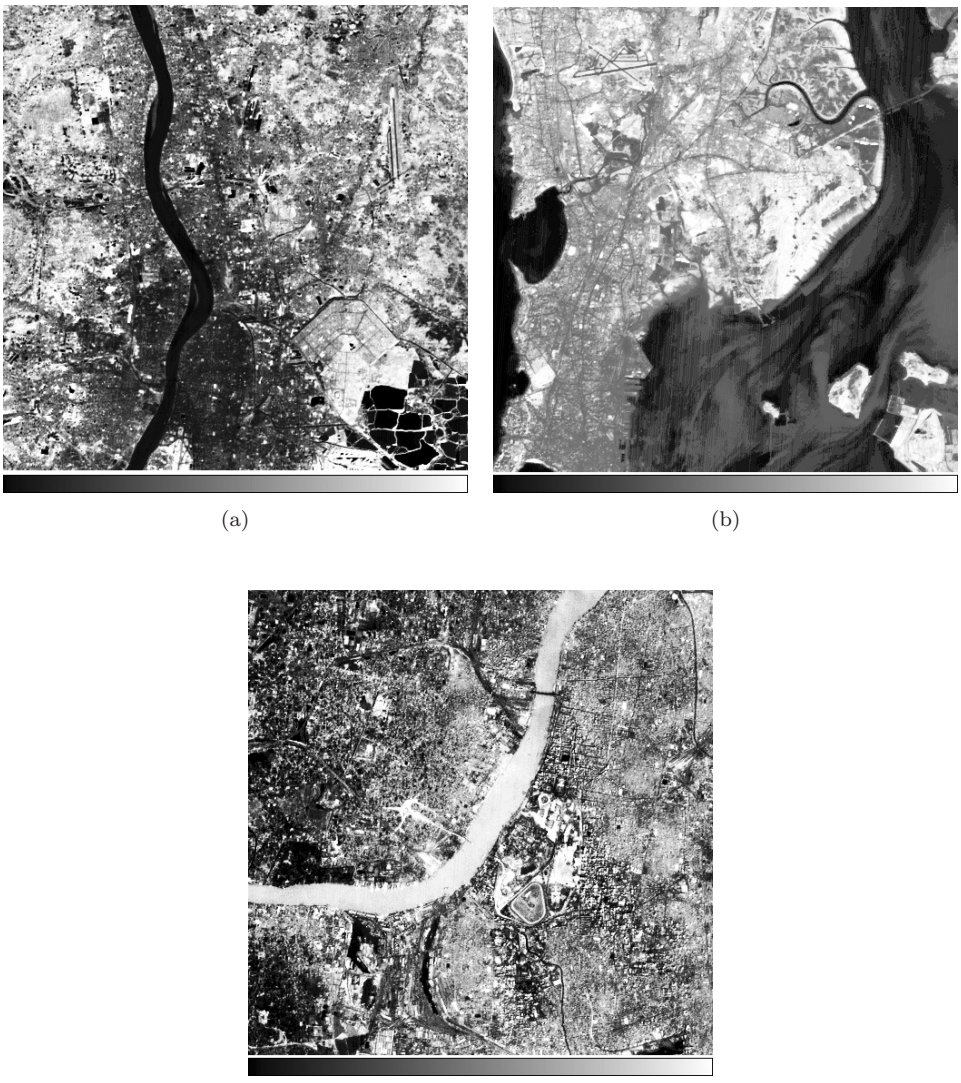
MLP is used in the present investigation to classify the land covers of remote sensing images. The performance of this classifier is evaluated on three images and the results are shown in tabular forms. Although we tested the algorithms on various wavelets, we have shown the performance of three better performing wavelets. They are: Biorthogonal 3.3 (Bior3.3) and Biorthogonal 3.5 (Bior3.5), and Daubechies 3 (Db3).[40] Classified images with Bior3.3 wavelet are shown in Figs. 6–8. In the present study we have used a two-level decomposition of the WT as the complexity increases proportionally with the level of decomposition with insignificant increase in the performance.

5.1. *Description of images*

5.1.1. *IRS-1A images*

Three different remote sensing images (size 512 × 512) are used for the simulation study of the proposed WF based classification scheme: two images are from Indian Remote Sensing Satellite 1A (IRS-1A) and one from SPOT (Systeme Pour d'Observation de la Terre) satellite. Due to poor illumination, the actual classes present in the input images are not visible clearly. So we have presented enhanced images in Fig. 5, which highlight the different land cover regions properly. However, the algorithm is implemented on actual (original) images.

The IRS-1A images [shown in Figs. 5(a) and 5(b)] were obtained from Indian Remote Sensing Satellite.[54] We have used the images taken from the Linear Imaging Self Scanner (LISS-II). LISS-II which 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



(a) (b)

(c)

Fig. 5. (a) and (b) Calcutta and Bombay map of IRS-1A image (band-4); and (c) Calcutta map of SPOT image (band-3).

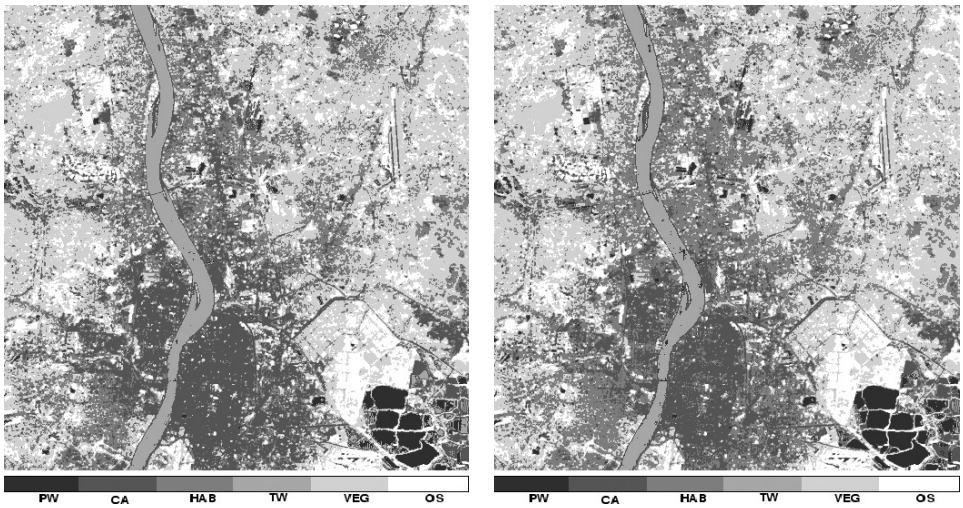
in Fig. 5(a) cover an area around the city of Calcutta in the near infrared band having six major land cover classes. These are *pure water* (PW), *turbid water* (TW), *concrete area* (CA), *habitation* (HAB), *vegetation* (VEG) and *open spaces* (OS). Fig. 5(b) shows a part of the Bombay city in the near infrared band. The elongated city area is surrounded by the *Arabian Sea*. The total region of the Bombay image can be classified into five major classes, namely *water* (W), *concrete area* (CA), *habitation* (HAB), *vegetation* (VEG) and *open spaces* (OS).

5.1.2. *SPOT image*

The SPOT image (enhanced) shown in Fig. 5(c) is obtained from SPOT satellite,[39] which carries an imaging device referred to as 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. This image has a higher spatial resolution of 20 m $\times$ 20 m. We have considered six different classes for the land cover classification of the SPOT image. These are *pure water* (PW), *river or turbid water* (TW), *concrete area* (CA), *habitation* (HAB), *vegetation* (VEG) and *open spaces* (OS).

5.2. *Classification of IRS-1A Calcutta image*

The classified (IRS-1A Calcutta) images obtained using MLP (original features) and MLP with Bior3.3 wavelet are only shown in Figs. 6(a) and 6(b), respectively as MLP performed well with Bior3.3 wavelet compared to other wavelets. From the visualization point of view, it is clear from the figures that the proposed NW classifier performed better in classifying the land covers (i.e. segregating different areas) compared to its corresponding neural version. Various objects in the IRS-1A Calcutta image are clearly identified in the classified images. For example, as shown in Fig. 6(b), we see that the *Hoogle* (*Ganges*) river situated in the middle of the image separating the image approximately into two halves, belongs to TW class. The *pure* or *fishery water* (PW class) is easily identified in the classified image. The other classes like CA, HAB, VEG and OS are also clearly visible. Objects like *Airport Runways, Saltlake Area, Saltlake Stadium, Vivekananda Bridge, Howrah*



(a) (b)

Fig. 6. Classified IRS-1A Calcutta image by (a) MLP (with original features); and (b) WT (Bior3.3)+ MLP classification method.

Bridge are also distinctly separable in the classified image using the proposed NW based classification scheme. The above mentioned objects are more or less visible in case of the classified images obtained by only MLP (with original features) as shown in Fig. 6(a). With the use of WT based features, the classes became more separated and well identified [Fig. 6(b)]. Also, a concrete distinction between various classes obtained by the classifier with different wavelets are justified with the evaluation of quantitative indices rather than only visualizing the regions.

Two quantitative indices, namely β and DB, as discussed in Sec. 4, have been used to justify the above findings. With β index, as discussed in the previous section, for a fixed number of classes, the higher the homogeneity behavior within the class, the higher the β value. Table 1 depicts the results of β. As expected, the β value is the highest for the training data, i.e. 9.4212 for IRS-1A Calcutta image. With MLP classification method, the β value is 7.1487. The β value is increased from 7.1487 to 7.7124, 7.7586 and 7.1678 for the wavelet feature based MLP classifier with Db3, Bior3.3, and Bior3.5 wavelets, respectively. From Table 1, it is clear that the classification result with Bior3.3 wavelet is providing the highest β value compared to others. As a whole, we can establish the following β relation in terms of the performance of the classifier with different wavelets

$$\beta_{\text{Bior3.3}} > \beta_{\text{Db3}} > \beta_{\text{Bior3.5}}.$$

From this relation it is observed that the performance of the classifier is better with Bior3.3 wavelet compared to other wavelets.

In the present investigation the performance is also compared with the percentage of accuracy (PA) achieved in the classification process using MLP and proposed WF based MLP classification methods. PA is calculated on the basis of β value obtained with training data. It is evident that the β value with training data is the highest. We assumed that the β value for training data is equivalent to 100% accuracy of the classifiers and the accuracy obtained with any of the classifiers can be compared on the basis of its β value. Thus PA can be expressed as

$$PA_\beta = \left(\frac{\beta_{cl}}{\beta_{tr}}\right) \times 100, \tag{5.1}$$

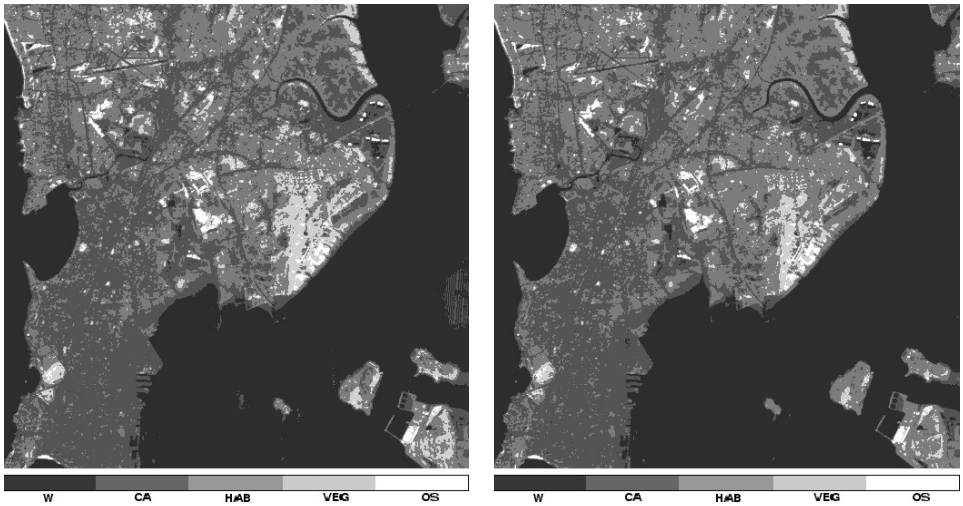
where β_{cl} and β_{tr} are the β values obtained from the classified image and from the training data, respectively. Using this expression, we evaluated the PA of MLP and the proposed NW classification methods. A PA of 75.87 and 82.35 are achieved with the use of MLP (with original features) and proposed NW classification methods, respectively. This shows that nearly **7%** increase in classification accuracy is achieved with the proposed classification method compared to original spectral feature based classifier.

Similar to the β index comparison, the DB index also supported the superiority of the proposed WF based NW classification method. The DB value using MLP for IRS-1A Calcutta image is 0.9347. This value is better (0.6813) with the Bior3.3 wavelet based features. Considering all cases we can infer that the combination

of MLP and Bior3.3 wavelet is outperforming the other wavelet and MLP (with original features) based methods.

5.3. *Classification of IRS-1A Bombay image*

In case of IRS-1A Bombay [Fig. 7(b)] image, the proposed MLP based NW method (with Bior 3.3 wavelet) also detected the regions like *Dockyard*, *Butcher Island*, *Elephanta Cave* and *Island* and *Santa Cruz Airport* in crisp and homogeneous way compared to the MLP method [original spectral feature only, Fig. 7(a)]. For other wavelets also similar conclusions can be made. Just one image is displayed for typical illustration. However, their quantitative index values are shown in Tables 1 and 2. As mentioned, the two classified images in Figs. 7(a) and 7(b) show some visual differences. However, a better comparison can be made from the quantitative indices like β and DB. Table 1 depicts the β values. The β, for the Bombay image with training data is found to be 21.4783, which is reduced to 17.6162 using MLP with original features. From the tables we notice that the β value for the proposed NW classifier with Bior 3.3 wavelet is 18.3815, which is again better than the values 18.3300 and 17.7232 obtained with the wavelets Db3 and Bior3.5, respectively. The same relations on β values are maintained for IRS-1A Bombay image as in the case of Calcutta image. The performance of the proposed classifier is also tested in terms of PA. It is found that the PA values are 85.58 and 82.01 using proposed WF and MLP (with original feature), respectively. The comparison with PA revealed the same remarks for the proposed classifiers.



(a) (b)

Fig. 7. Classified IRS-1A Bombay image by (a) MLP (with original features); and (b) WT (Bior3.3) + MLP classification method.

Table 1. Comparison of β and PA values for different neuro-wavelet classification methods.

Classifier	Wavelet	IRS Calcutta image		IRS Bombay image		SPOT Calcutta image	
		β	$PA_\beta(\%)$	β	$PA_\beta(\%)$	β	$PA_\beta(\%)$
TP*	—	9.4212	100	21.4783	100	9.3343	100
MLP	—	7.1487	75.87	17.6162	82.01	7.0341	75.35
WT + MLP	Db3	7.7124	81.86	18.3300	85.34	7.6121	81.54
WT + MLP	**Bior3.3**	**7.7586**	**82.35**	**18.3815**	**85.58**	**7.7143**	**82.64**
WT + MLP	Bior3.5	7.1678	76.08	17.7232	82.51	7.2013	77.14

Note: *TP = Training patterns.

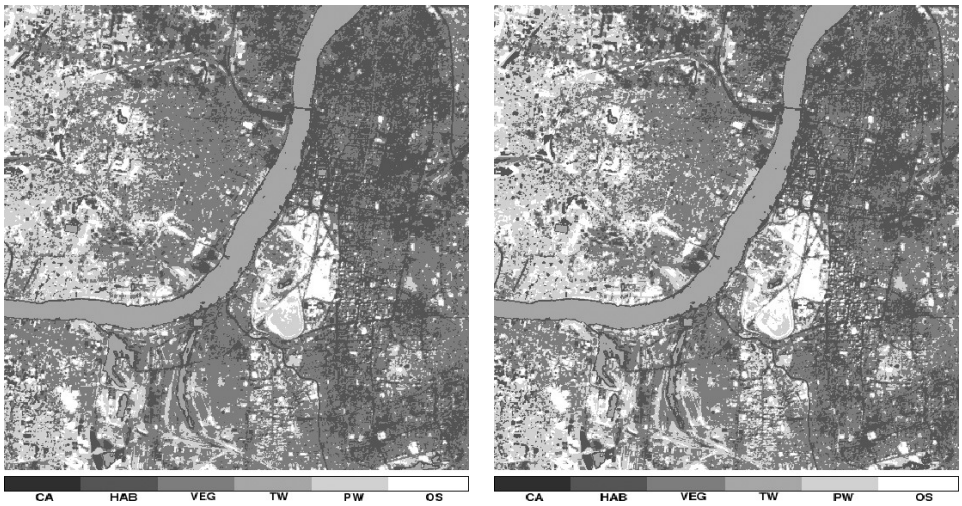
Table 2. Comparison of DB value for neuro-wavelet classification method.

Classifier	IRS Calcutta image	IRS Bombay image	SPOT Calcutta image
MLP	0.9347	0.9246	2.3105
MLP+Bior3.3 (wavelet)	**0.6813**	**0.6785**	**1.4947**

Similarly, the DB index shown in Table 2 also indicates the superiority of the proposed method. The Bior3.3 wavelet based MLP provided promising classification results over others. The DB value for this method is 0.6785, which is the lowest and justifies the efficiency of the proposed method in classifying the land covers present in the IRS-1A Bombay image.

5.4. *Classification of SPOT Calcutta image*

For SPOT Calcutta image, the classified regions of the images are shown in Fig. 8(a) for MLP (original spectral features only) and Fig. 8(b) for the proposed NW classifier, i.e. WF with Bior3.3 wavelet, only as this was performing better compared to others. From the figures, it is observed that there is a clear separation of different classes and some known regions like *Race Course*, *Howrah Bridge* (*Setu*), *Talis Nala* (*Canal*), *Beleghata Nanal*, *Khiderpore Dock* and *Garden Reach Lake* by the proposed method. It is also evident that the proposed method produced a well structured and proper shaped regions compared to the MLP with original spectral features. However, a better performance comparison with the help of β value can be seen from Table 1. The β value for the training data set is 9.3343, and it is 7.0341 when the test patterns are classified with MLP. The proposed NW classifier is providing a higher β value compared to its corresponding original spectral feature based version. A further improvement of the MLP classifier with WF is observed with Bior3.3 wavelet. The β relation, in the classification of SPOT Calcutta image, are also observed to be similar to the case of IRS-1A Calcutta and IRS-1A Bombay images. The comparative analysis in terms of PA for both the classifiers are made and observed that the proposed classifier provided 82.64 accuracy, which is more



(a) (b)

Fig. 8. Classified SPOT Calcutta image by (a) MLP (with original features), and (b) WT (Bior3.3) + MLP classification method.

compared to the PA value of 75.35 obtained with the MLP. Thus, there is an increase of nearly 7% of PA value by the proposed classifier compared to MLP with original spectral features only.

Like the β index, the DB values also corroborate the earlier findings. The comparative DB values for the proposed and MLP with original feature based classification methods are shown in Table 2, which strengthens the previous findings.

6. Conclusion

In the present article, we have proposed a WF based neural (MLP) approach for classification of multispectral remote sensing images. The proposed scheme tried to explore the possible advantages of using wavelet transform (WT) as a preprocessor for the neural classifier. The WT is used to extract features from the original patterns. The extracted features acquire information of the pixel along with its neighbors both in spatial and spectral domains because of the inherent characteristic of the WT which makes the next stage of classification more efficient compared to that of without preprocessing. We have used a two-level decomposition of WT, as complexity increases proportionally with the level of decomposition with insignificant increase in performance.

The improvement in performance of the proposed neuro-wavelet classification scheme is verified from the results obtained from classification of three remote sensing images. The β values (shown in Table 1) indicate the classification accuracy and support the claim of better visual separation for the identification of various

known regions in a significantly improved way. We modified the original β index to compute the percentage accuracy based on $\beta_{training}$, which provided a better indicator for assessing the classifiers. The PA_β value revealed that there is an increase of accuracy (*nearly* 7% *for IRS-1A Calcutta and SPOT images and around* 4% *in case of IRS-Bombay*) using neuro-wavelet classifier with Bior3.3 wavelet, which is reasonably significant. The DB index value also supported the above findings. Different wavelets are used in the preprocessing stage of the proposed scheme. Bior3.3 wavelet is found to be more appropriate than others for the present investigation. Also, it is observed that the different structures of the classified regions obtained with the proposed method are more crisp and well shaped. Thus, in conclusion, we can say that for the present set of images, proposed neuro-wavelet classifier with Bior3.3 wavelet is well suited for the detection of objects in remote sensing images.

Future studies will include neural networks with curvelet, countourlet and higher dimensional wavelets for classification of land covers in multispectral remote sensing images. The utility of other neural network models like radial basis function network will also be explored.

Acknowledgments

The authors would like to thank the Department of Science and Technology, Government of India, for sponsoring the project titled "Advanced Techniques for Remote Sensing Image Processing".

References

1. B. Tso and P. M. Mather, *Classification Methods for Remotely Sensed Data* (Taylor and Francis, London, 2001).
2. R. M. Haralick and K. S. Shanmugam, Combined spectral and spatial processing of ERTS imagery data, *Remote Sens. Env.* **3** (1974) 3–13.
3. S. Tuominen and A. Pekkarinen, Performance of different spectral and textural aerial photograph features in multi-source forest inventory, *Rem. Sens. Env.* **94** (2005) 256–268.
4. E. H. H. Shih and R. A. Schowengerdt, Classification of arid geomorphic surfaces using spectral and textural features, *Photogramm. Eng. Rem. Sens.* **49** (1983) 337–347.
5. X. Yuan, D. King and J. Vlcek, Sugar maple decline assessment based on spectral and textural analysis of multispectral aerial videography, *Remote Sens. Env.* **37** (1991) 47–54.
6. M. Chica-Olmo and F. Abarca-Hernndez, Computing geostatistical image texture for remotely sensed data classification, *Comp. Geosci.* **26** (2000) 373–383.
7. G. R. Cross and A. K. Jain, Markov random field texture models, *IEEE Trans. Pattern Anal. Mach. Intell.* **5** (1983) 25–39.
8. R. Chellappa and S. Chatterjee, Classification of textures using Gaussian Markov random fields, *IEEE Trans. Acoust. Speech. Signal Process.* **4** (1986) 959–963.

9. F. S. Cohen, Z. Fan and M. A. Patel, Classification of rotation and scaled textured images using Gaussian Markov random field models, *IEEE Trans. Pattern Anal. Mach. Intell.* **13** (1991) 192–202.
10. H. Derin and H. Elliot, Modeling and segmentation of noisy and textured images using Gibbs random fields, *IEEE Trans. Pattern Anal. Mach. Intell.* **9** (1987) 39–59.
11. M. Unser, Local linear transforms for texture measurements, *Signal Process.* **11** (1986) 61–79.
12. A. Jouan and Y. Allard, Land use mapping with evidential fusion of features extracted from polarimetric synthetic aperture radar and hyperspectral imagery, *Inform. Fus.* **5** (2004) 251–267.
13. S. Mallat, *A Wavelet Tour of Signal Processing* (Academic Press, 1999).
14. G. Strang and T. Nguyen, *Wavelets and Filter Banks* (Wellesley College, 1996).
15. M. Unser, Texture classification and segmentation using wavelet frames, *IEEE Trans. Image Process.* **4** (1995) 1549–1560.
16. T. Chang and C. C. J. Kuo, Texture analysis and classification with tree-structured wavelet transform, *IEEE Trans. Image Process.* **2** (1993) 429–440.
17. M. Unser and M. Eden, Multiresolution feature extraction and selection for texture segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* **11** (1989) 717–728.
18. G. V. de Wouwer, P. Schenders and D. V. Dyek, Statistical texture characterization from discrete wavelet representation, *IEEE Trans. Image Process.* **8** (1999) 592–598.
19. A. Abbate, P. Das and C. M. Decusatis, *Wavelets and Subbands: Fundamentals and Applications* (Springer, 2001).
20. S. Arivazhagan and L. Ganesan, Texture classification using wavelet transform, *Pattern Recog. Lett.* **24** (2003) 1513–1521.
21. H. H. Szu, J. L. Moigne, N. S. Netanyahu and C. C. Hsu, Integration of local texture information in the automatic classification of landsat images, *SPIE Proc.* **3078** (1997) 116–127.
22. J. Yu and M. Ekstrom, Multispectral image classification using wavelets: A simulation study, *Pattern Recog.* **36** (2003) 889–898.
23. S. K. Meher, B. Uma Shankar and A. Ghosh, Multispectral remote sensing image classification using wavelet based features, in *Soft Computing in Image Processing*: *Recent Advances*, eds. M. Nachtegael, D. Vin der Weken, E. E. Kerre and W. Philips (Springer, 2006), pp. 3–34.
24. S. K. Meher, B. Uma Shankar and A. Ghosh, Remote sensing image classification: A wavelet-neuro-fuzzy approach, in *Proceedings of International Conference on Advances in Pattern Recognition* (World Scientific, 2007), pp. 231–236.
25. B. Uma Shankar, S. K. Meher and A. Ghosh, Neuro-wavelet classifier for remote sensing image classification, in *Proceedings of International Conference on Computing*: *Theory and Applications* (IEEE Computer Society Press, 2007), pp. 711–715.
26. B. K. Alpert, Wavelets and other bases for fast numerical linear algebra, in *Wavelets*: *A Tutorial in Theory and Applications*, ed. C. K. Chui (Academic Press, 1992), pp. 181–216.
27. V. Strela, P. Heller, G. Strang, P. Topiwala and C. Heil, The application of multiwavelet filter banks to image processing, *IEEE Trans. Image Process.* **8** (1999) 548–563.
28. H. Soltanian-Zadeh, F. Rafiee-Rad and S. Pourabdollah-Nejad, Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms, *Pattern Recog.* **37** (2004) 1973–1986.
29. S. Haykin, *Neural Networks: A Comprehensive Foundation* (Prentice Hall, 1997).

30. P. F. Baldi and K. Homik, Learning in linear neural networks: A survey, *IEEE Trans. Neural Network.* **6** (1995) 837–858.
31. G. P. Zhang, Neural networks for classification: A survey, *IEEE Trans. Syst. Man Cyb.* **30** (2000) 451–462.
32. S. Marinai, M. Gori and G. Soda, Artificial neural networks for document analysis and recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* **27** (2005) 23–35.
33. J. Bendiktsson, P. Swain and O. Ersoy, Neural network approaches versus statistical methods on classification of multisource remote sensing data, *IEEE Trans. Geosci. Rem. Sens.* **28** (1990) 540–552.
34. H. Bichof, W. Schneider and A. Pinz, Multishpectral classification of landsat image using neural networks, *IEEE Trans. Geosci. Rem. Sens.* **30** (1992) 482–490.
35. P. D. Heermann and N. Khozenie, Classification of multispectral remote sensing data using a back-propagation neural network, *IEEE Trans. Geosci. Rem. Sens.* **30** (1992) 81–88.
36. T. Kavzoglu and P. M. Mather, The use of backpropagating artificial neural networks in land cover classification, *Int. J. Rem. Sens.* **24** (2003) 4907–4938.
37. F. S. Erbek, C. O. Zkan and M. Taberner, Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities, *Int. J. Rem. Sens.* **25** (2004) 1733–1748.
38. D. E. Rumelhart, G. E. Hinton and R. J. Williams, Learning representations by back-propagating errors, *Nature* **323** (1986) 533–536.
39. J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction* (Springer-Verlag, 1999).
40. I. Daubechies, *Ten Lectures on Wavelets* (SIAM, 1995).
41. H. Liu and H. Motoda, *Feature Extraction, Construction and Selection*: *A Data Mining Perspective* (Kluwer Academic Publishers Norwell, 1998).
42. J. Mazzaferri, S. Ledesma and C. Iemmi, Multiple feature extraction by using simultaneous wavelet transforms, *J. Opt. A*: *Pure Appl. Opt.* **5** (2003) 425–431.
43. S. Soltani, P. Simard and D. Boichu, Estimation of the self-similarity parameter using the wavelet transform, *Signal Process.* **84** (2004) 117–123.
44. F. A. Mujica, J. P. Leduc, R. Murenzi and M. J. T. Smith, A new motion parameter estimation algorithm based on the continuous wavelet transform, *IEEE Trans. Image Process.* **9** (2000) 873–888.
45. P. K. Topiwala, *Wavelet Image and Video Compression* (Springer, 2000).
46. D. Taubman and M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice* (Springer, 2001).
47. M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding* (Prentice Hall, 1995).
48. R. C. Gonzalez and R. E. Woods, *Digital Image Processing* (Prentice Hall, 2002).
49. J. Cohen, A coefficient of aggrement for nominal scale, *Educ. Psychol. Meas.* **20** (1960) 37–46.
50. S. K. Pal, A. Ghosh and B. U. Shankar, Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation, *Int. J. Rem. Sens.* **21** (2000) 2269–2300.
51. M. Acharyya, R. K. De and M. K. Kundu, Segmentation of remotely sensed images using wavelet features and their evaluation in soft computing framework, *IEEE Trans. Geosci. Rem. Sens.* **41** (2003) 2900–2905.

52. P. Mitra, B. U. Shankar and S. K. Pal, Segmentation of multispectral remote sensing images using active support vector machines, *Pattern Recog. Lett.* **25** (2004) 1067–1074.
53. D. L. Davies and D. W. Bouldin, A cluster separation measure, *IEEE Trans. Pattern Anal. Mach. Intell.* **1** (1979) 224–227.
54. "IRS data users hand book", Technical Report, IRS/NRSA/NDC/HB-02/89 (1989).