

Opening the black box of neural networks for remote sensing image classification

F. QIU

Program in Geographic Information Sciences, University of Texas at Dallas,
Richardson, TX 75083, USA; e-mail: ffqiu@utdallas.edu

and J. R. JENSEN

Department of Geography, University of South Carolina, Columbia,
SC 29205, USA

(Received 18 July 2001; in final form 2 June 2003)

Abstract. Neural networks, which make no assumption about data distribution, have achieved improved image classification results compared to traditional methods. Unfortunately, a neural network is generally perceived as being a ‘black box’. It is extremely difficult to document how specific classification decisions are reached. Fuzzy systems, on the other hand, have the capability to represent classification decisions explicitly in the form of fuzzy ‘if-then’ rules. However, the construction of a knowledge base, especially the fine-tuning of the fuzzy set parameters of the fuzzy rules in a fuzzy expert system, is a tedious and subjective process. This research has developed a new, improved neuro-fuzzy image classification system based on the synergism between neural networks and fuzzy expert systems. It incorporates the best of both technologies and compensates for the shortcomings of each. The learning algorithms of neural networks developed here are used to automate the derivation of fuzzy set parameters for the fuzzy ‘if-then’ rules in a fuzzy expert system. The rules obtained, in symbolic form, facilitate the understanding of the neural network based image classification system. In addition, the image classification accuracy obtained from the improved neuro-fuzzy system was significantly superior to those of the back-propagation based neural network and the maximum likelihood approaches.

1. Introduction

Data acquired by remote sensing systems can be used to systematically extract a variety of fundamental biophysical and land use/land cover information over large geographic areas if processed with the appropriate analytical tools (Jensen 2000). The recent availability of high spatial resolution imagery (e.g. Space Imaging Inc. IKONOS 1 m × 1 m panchromatic and 4 m × 4 m multispectral) and hyperspectral data (e.g. NASA’s Hyperion sensor system) may be very useful for extracting land use and land cover information (Jensen and Cowen 1999). However, the higher spatial resolution imagery now contains even more high frequency information. Similarly, the greater number of bands introduces additional redundancy in information content and vastly complicates the task of image processing. In effect,

this increased information expands the volume of feature space occupied by a particular land use/land cover class and may cause many classes to overlap each other. Consequently, classification can become more complex, demanding the development of more powerful image classification tools. The same well established statistical approaches that provide acceptable classification with relatively low spatial and spectral resolution imagery, fail to provide satisfactory results when applied to improved remote sensor data (Jensen 1996, Roli *et al.* 1996).

Many analysis techniques have been proposed as alternatives to traditional classification tools to process complex remote sensing data. Artificial neural networks (ANNs) and fuzzy logic systems are two of the most innovative techniques that have shown great potential in analysing and making use of improved remote sensor data.

1.1. *Neural network systems*

Artificial neural networks have been employed to process multispectral remote sensing images and have achieved improved accuracy compared to those obtained from traditional statistical methods (Heermann and Khazenie 1992, Civco 1993, Foody *et al.* 1995, Gong 1996, Benediktsson and Sveinsson 1997). These successes are underpinned by many salient characteristics of neural networks. A single neuron simulates the computation of a multivariate linear regression model (Hewitson and Crane 1994, Jensen *et al.* 1999), making no *a priori* assumptions of normal and linear data distribution due to its operation in a non-parametric fashion (Lippmann 1987, Haykin 1994, Foody *et al.* 1995, Lloyd 1996). Neurons with nonlinear activation functions are arranged in layers and act like a set of piece-wise nonlinear simulators (Hornik *et al.* 1989, Haykin 1994). Neural networks are able to learn from existing examples adaptively, which makes the classification objective (Hagan *et al.* 1996). At the same time, various noise information inevitably included in the examples supply a trained neural network with the ability to generalize, which makes neural networks robust solutions in the presence of previously unseen, incomplete or imprecise data (Fausett 1994).

Despite the excellent performance of neural networks in image classification, it is difficult to provide a comprehensible explanation of the process through which a given output has been obtained from a neural network (Andrews *et al.* 1995). Through network training, the knowledge of image classification can be derived and stored implicitly in numerical forms as synaptic weights in the network. Unlike the parameters of regression functions, these weights have no obvious meaning in most cases. It is difficult, if not impossible, to interpret these weights due to their complex nature (Lein 1997). Therefore, a neural network is often accused of being a black box, which hides the relation between inputs and outputs in the weights of the neurons of its 'hidden' layers (Benitez *et al.* 1997). As a result, we cannot gain any understanding of the problem at hand due to the lack of an explanatory capability to provide insight into the characteristics of the dataset. For the same reason, it is also impossible to incorporate human expertise to simplify, accelerate or improve the performance of image classification – a neural network always has to learn from scratch (Nauck *et al.* 1997). For neural networks to be widely applied to complex remote sensing image classification tasks, an explanation capability should be an integral part of the functionality of a trained neural network (Andrews *et al.* 1995).

1.2. Fuzzy systems

A fuzzy system, or fuzzy expert system, on the other hand, is able to represent image classification decisions explicitly in the form of declarative fuzzy 'if-then' rules. Fuzzy systems are an extension of traditional expert systems. Like traditional expert systems, rules in fuzzy systems are extracted directly and mainly from one or many knowledgeable expert(s) of a domain area and then used in an algorithmic, structured manner to answer questions of this restricted domain application. Unlike traditional expert systems, the 'if-then' rules in fuzzy systems utilize the concept of a fuzzy set (Zadeh 1965), an extension and generalization of the classic Boolean set. Fuzzy sets allow the assignment of partial and multiple valued memberships. Fuzzy systems deliberately make use of vague, imprecise or uncertain information to generate simpler, more suitable models that are easier to handle and more familiar to human thinking (Klir and Yuan 1995). Therefore, fuzzy sets are suitable for representing a mixed pixel, which is defined as one that records a mixture of radiant flux from more than one ground biophysical phenomenon within the instantaneous-field-of-view of a sensor system (Wang 1990, Jensen 1996). With linguistic terms, fuzzy 'if-then' rules establish a direct mapping from the antecedent part (i.e. input variables) to the consequent part (i.e. output variables) of each rule and thus provide an understanding of the processes involved. Human expert knowledge is the main source of these fuzzy rules. Thus, it is also possible to improve system performance by adding new rules, removing defective rules and/or update existing rules in the knowledge base.

The application of fuzzy expert systems to classify remote sensing imagery has not been widely reported except by Binaghi *et al.* (1997). This situation is due primarily to the limitation of fuzzy systems in the acquisition of expert knowledge. Fuzzy expert systems rely solely on the encapsulation of human expertise to derive fuzzy 'if-then' rules and to specify fuzzy set parameters. The lack of a mechanism to learn from examples makes the acquisition of knowledge a tedious and subjective process (Openshaw and Openshaw 1997). Even if experts being interviewed are aware of the underlying intuitive rules of thumb that they have used and are able to articulate these rules, the construction of a fuzzy system needs the definition of suitable fuzzy sets to represent the linguistic terms or fuzzy numbers used in the fuzzy rules. In complex fuzzy systems, manual determination and optimization of fuzzy membership parameters is virtually impossible (Abe and Lan 1996). These problems have prevented fuzzy expert systems from being widely applied in routine remote sensing image processing. It is desirable that knowledge automation capability be incorporated into existing fuzzy systems, so that the benefits promised by fuzzy set theory can be made available to the image classification task (Medsker 1994).

1.3. Integrated neural network and fuzzy systems

Both neural network and fuzzy systems are attempts to bring 'human intelligence' into traditional data processing tasks, although the implementation is approached from different perspectives (Kulkarni and Lulla 1999). As stand-alone systems, neural networks and fuzzy systems exhibit unique features and fundamental limitations as well. Consequently, each technology has the capability to solve one aspect of the problem, but neither by itself can provide a total solution.

Recently, there has been research interest in synthesizing and/or integrating neural networks and fuzzy systems. The examination of their assets and liabilities

reveals an interesting and exciting fact: the two technologies are coincidentally complementary with each other from a functional point of view. The benefits possessed by one technology happen to be shortcomings of the other (Medsker 1994, Kulkarni and Lulla 1999). Therefore, if these two approaches are combined, one technology can provide the capabilities not available in the other. The integration of neural networks and fuzzy systems is often known as neuro-fuzzy systems in artificial intelligence (Nauck *et al.* 1997). A neuro-fuzzy system is basically a fuzzy system that uses a learning algorithm derived from or inspired by neural network theory to determine its parameters based on sample data (Nauck *et al.* 1997). A neuro-fuzzy system usually delivers a more powerful solution than its individual components. In the subsequent subsections, three existing neuro-fuzzy systems were evaluated and their respective advantages and limitations for remote sensing image classification were analysed.

1.3.1. Back-propagation based neuro-fuzzy system

The back-propagation based neuro-fuzzy system was derived by adopting the interpretation of the popular multi-layer perceptron (MLP) model by Benitez *et al.* (1997). This system used a unique three-layer topological structure, where the number of neurons in the input layer corresponds to the number of input variables, and the number of neurons in the hidden layer is equal to that of output variables. The output layer contains, however, only one neuron. In the hidden layer, a continuous and nonlinear sigmoid function is utilized as the activation function, but in the output layer an identity function is employed. The sigmoid function used in a neural network as an activation function can also be used in a fuzzy system as a membership function. Benitez *et al.* (1997) utilized the concept of f -duality to facilitate the extraction of the fuzzy 'if-then' rule from this neural network. It was discovered that the f_A -dual of the summation operator \sum is an *interactive-or* operator (*i-or*) * of n arguments. As a result, the output of a hidden neuron y can be interpreted as a fuzzy 'if-then' rule:

$$\text{if } x_1 \text{ is } A_{i1} \text{ i-or } x_2 \text{ is } A_{i2} \text{ i-or } \dots \text{ i-or } x_j \text{ is } A_{ij} \dots \text{ i-or } x_n \text{ is } A_{in} \text{ then } y = a_i \quad (1)$$

where A_{ij} is the fuzzy number defined by a sigmoid function f_A relating j th input variable x with i th target information class a_i and n is the total number of input variables.

The development of such a neuro-fuzzy system is straightforward because the system can be readily implemented by employing an existing MLP neural network with the specified topological structure and activation functions. No modification of the neural network is required and the interpretation of the weights is independent of the training algorithm used in the MLP network, although a back-propagation algorithm is used here.

Unfortunately, the use of the back-propagation based neuro-fuzzy system to extract fuzzy 'if-then' rules for image classification encounters several problems. A sigmoid function is an open fuzzy membership function and can be used to represent a fuzzy number, such as 'approximately >40 ' (figure 1(a)). However, normally two such functions (such as 'approximately >20 ' and 'approximately <60 ') are needed to represent a fuzzy interval number (such as 'approximately 40') (figure 1(b)). The fuzzy rules extracted from this system do not include two functions that stand for a fuzzy interval number. The fuzzy 'if-then' rules extracted are stated with regard to the relationship between input variables and the output of

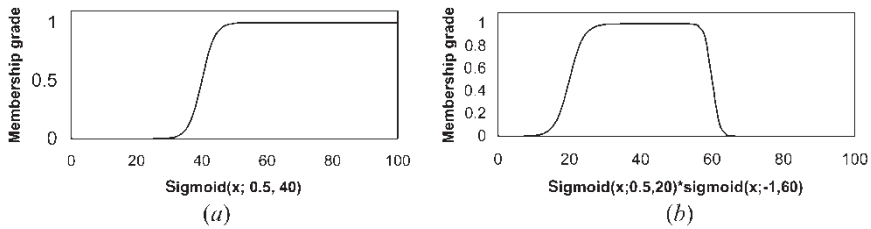


Figure 1. Sigmoid membership functions representing a fuzzy number. (a) One sigmoid function is used to represent fuzzy number 'approximately >40 '. (b) Two sigmoid functions are needed to represent fuzzy number 'approximately 40'.

the hidden neurons. This complicates the syntactical understanding of the system because the rules fail to establish a direct link from the input variables to the target information class. The employment of a complicated fuzzy aggregation (*interactive-or*) operator makes the interpretation of the extracted rules more semantically complex. Benitez *et al.* (1997) reported that the fuzzy rules generated by a MLP network may contain out-of-range fuzzy numbers for the training dataset. This situation occurred when the network was applied to remote sensing data and caused the membership parameters of the fuzzy rules to fail to fall within the range of 0 to 255 in the remote sensing data. Therefore, the rules generated by the back-propagation based neuro-fuzzy system were found to not be helpful in obtaining a clear understanding of the image classification process.

1.3.2. Fuzzy self-organizing map based neuro-fuzzy system

This neuro-fuzzy system was derived from the fuzzy self-organizing map (FSOM) proposed by Vuorimaa (1994). FSOM is built on the basis of the fuzzification of the well known Kohonen self-organizing map (SOM) neural network (Kohonen 1989) using a triangular membership function. Traditional self-organizing maps use the n -dimensional Euclidean distance between an input vector and the corresponding weight vectors to determine a 'winner' output neuron based on the minimum distance (i.e. the one that is closest to the input vector). The learning rule of a self-organizing map has the capability to shift the weight vector of a 'winner' neuron to the centroid of a data cluster.

Vuorimaa's (1994) fuzzy version of the SOM first employed the traditional learning SOM rules to initialize the centroid of each data cluster. Then the new model fuzzified the traditional SOM with a set of fuzzy 'if-then' rules defined by a triangular membership function, whose centre corner was specified by the calculated centroid of each data cluster. In order to determine the overall membership grade of each rule, the fuzzy SOM (FSOM) uses the standard fuzzy intersection operator (i.e. the minimum function) to obtain the output. The fuzzy self-organizing map assigns the 'winner' to the output neuron that has the highest degree of overall membership grade. The fuzzy set parameters (i.e. the centre and the left and right spreads of the triangular function) are then fine-tuned using a modified version of the learning vector quantization (LVQ) algorithm (Kohonen *et al.* 1996). The LVQ algorithm adjusts the fuzzy set parameters only if the input value falls within the overlapped area determined by the triangular membership functions for the 'winner' rule and the 'first runner-up' rule. The final output of the network is the weighted average of each rule's output and its corresponding membership grade, which produces a continuous value output.

Vuorimaa’s (1994) FSOM was designed to simulate continuously valued functions. The FSOM is a supervised learning neural network although the original Kohonen SOM is not. Provided with enough pairs of training examples, the network is able to model any continuous function to an arbitrary degree of accuracy. The replacement of Kohonen’s original model (Kohonen 1989) with a set of fuzzy ‘if-then’ rules that are defined by simple triangular fuzzy membership functions helps to understand the processing involved in the neural network.

When applied to image classification problems, however, the algorithm did not achieve convergence and failed to produce useful image classification results. This is because the output of a classification problem produces discrete categories rather than continuous values. The fine tuning of the fuzzy set parameters based on the modified LVQ requires the overlapping of the two most fired fuzzy rules, which are not always easy to obtain for image classification problems. As a consequence, the network did not achieve convergence because there is no guarantee that two adjacent triangular membership functions will overlap each other (figure 2(a)). In addition, the triangular membership function assumes a 0 membership grade if a pixel value in one of the bands is outside the range determined by the lower and upper limits of a triangular function. It is possible that at many places in the feature space, the brightness values of a pixel are not covered by any triangular function (figure 2(a)). Therefore, preliminary trials of this system yielded many unclassified pixels using standard fuzzy intersection functions. Although it may generate easily understood fuzzy rules, the FSOM resulted in an unacceptable image classification.

1.3.3. Fuzzy inference network (FIN) based neuro-fuzzy system

The fuzzy inference network (FIN) based neuro-fuzzy system was based on the work of Nomura and Miyoshi (1995). Similar to Vuorimaa’s (1994) system, this neuro-fuzzy system was based on the fuzzification of the Kohonen SOM (Kohonen 1989). Instead of using triangular functions, however, the FIN model uses Gaussian membership functions to fuzzify the traditional SOM. For each fuzzy ‘if-then’ rule, the system uses an algebraic product fuzzy intersection operator to calculate the overall membership grade. Similarly, the ‘winner’ neuron of the FSOM is the output neuron that produced the highest degree of overall membership grade. The FSOM updates only the weight parameters of the ‘winner’ neuron and the neurons within a specified neighbourhood of this neuron. For the centre parameter of the Gaussian function, the system adjusts its value using the traditional Kohonen (1989) SOM update rule. For the standard deviation parameter σ of the Gaussian function that was used to define the fuzzy set boundary, the system fine-tunes its

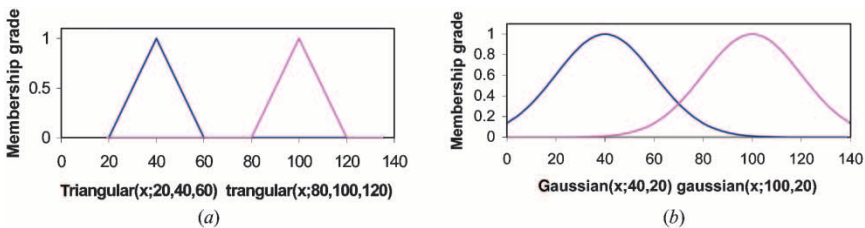


Figure 2. (a) Two triangular membership functions may not overlap each other, thus values between two triangular functions may have a membership grade of 0; (b) Two Gaussian membership functions always overlap with each other, thus values between two triangular functions could not have a membership grade of 0.

value using a different update rule:

$$\Delta\sigma_{ij} = 2\eta\sigma_{ij}\left((c_{ij} - x_j)^2 - \sigma_{ij}^2\right) \quad (2)$$

where η is the learning rate and x is the input, and c_{ij} is the centre parameter of the Gaussian function relating i th target information class with j th input variable x_j . Based on this fuzzified neural network, a fuzzy inference network (FIN) is established. The FSOM constitutes the 'if' part of the FIN. The 'then' part of the FIN is formed by weighted average of the outputs from the fuzzy 'if-then' rules, which are designed to produce continuous output.

Nomura and Miyoshi's (1995) FSOM possesses similar advantages to Vuorimaa's (1994) model. However, the use of Gaussian membership functions instead of triangular functions endows this system with yet another feature. Mathematically, any two one-dimensional Gaussian functions will intersect each other (figure 2(b)). Therefore, no pixel value in a remote sensor band will have a 0 membership grade. This solves the unclassified pixel problem suffered by Vuorimaa's (1994) FSOM.

Unfortunately, the fuzzy inference network also exhibits some drawbacks. First, the learning rule for the standard deviation parameter of the Gaussian function (2) is an update function that should be used to derive variance. Using remote sensing data within the original range of 0 and 255 will cause the system to crash. The square of a number that is greater than 1 will make the adjustment of the weight increase exponentially. It does not take many iterations before the adjustment becomes so huge that it exceeds the maximum double precision real number that a computer can represent. On the other hand, if raw values are scaled to the range between 0 and 1, the learning process becomes extremely slow. This is because the square of a decimal number between 0 and 1 produces a smaller decimal also in the range of 0 to 1. Sooner or later, the adjustment will become extremely small and approximate 0, thus greatly slowing the learning process. Furthermore, the employment of the algebraic product intersection operator tends to underestimate the final membership grade of a pixel. For example, if the fuzzy membership grades of a pixel for all three bands were 0.5, one would expect that the final membership grade would also be 0.5. However, the use of an algebraic product function produces an overall membership grade of 0.125 (i.e. $0.5 \times 0.5 \times 0.5$), which is much lower than that of the original individual band.

As a result, this neuro-fuzzy system is unstable in producing an accurate classification. Furthermore, the learning speed is extremely slow when using scaled remote sensing data, and generates unreasonable overall membership grades.

1.4. *The objective of this research*

The three neuro-fuzzy systems investigated exhibited certain advantages as well as some limitations. Applied to real world remote sensing data, none of the systems was able to produce an accurate image classification, extract easily understood fuzzy 'if-then' rules, and generate reasonable overall fuzzy membership at the same time. The objective of this research is to develop a neuro-fuzzy system specifically designed for remote sensing image classification based on incorporating the advantages of the aforementioned neuro-fuzzy systems while best minimizing their disadvantages. The development of this improved neuro-fuzzy system utilizes the learning algorithm of a neural network to automatically fine-tune the associated parameters of the fuzzy 'if-then' rules that it helps to form. The derived rules, in

symbolic form, are then used to facilitate the understanding of the neural network-based remote sensing image processing system. In this way, the benefits of neural networks and fuzzy systems can be brought together while their drawbacks are avoided. The result of this integration is that the black box of neural networks can be opened, and the knowledge acquisition process associated with fuzzy expert systems can be facilitated.

2. Methodology

The improved neuro-fuzzy system was developed using ERDAS Imagine’s Developer Toolkit. It was expected to produce an accurate image classification comparable to the back-propagation based system, extract understandable fuzzy ‘if-then’ rules as with the FIN based system, and generate a reasonable overall fuzzy membership like the FSOM based system.

The improved neural network developed was based on a fuzzified version of the learning vector quantization (LVQ) network (Kohonen *et al.* 1996), a supervised version of Kohonen’s SOM (Kohonen 1989). The fuzzified LVQ network consists of three layers: an input layer, a competitive layer and an output layer. The neurons in the input layer correspond to the input variables (i.e. the brightness values of a pixel for multiple bands). The number of neurons in the competitive layer was equal to or greater than that of the output classes (e.g. two or three times the number of output classes). In this way, it was possible to extract multiple rules for one class such that the multi-modal distribution of remote sensing data could be addressed. The number of neurons in the output layer was equivalent to that of the target information classes (i.e. land use/land cover types).

The neuro-fuzzy system constructed contains the fundamental ingredients of both a neural network and a fuzzy system. It consists of the following components (figure 3):

- 1. fuzzification interface,
- 2. fuzzy inference engine and knowledge base,

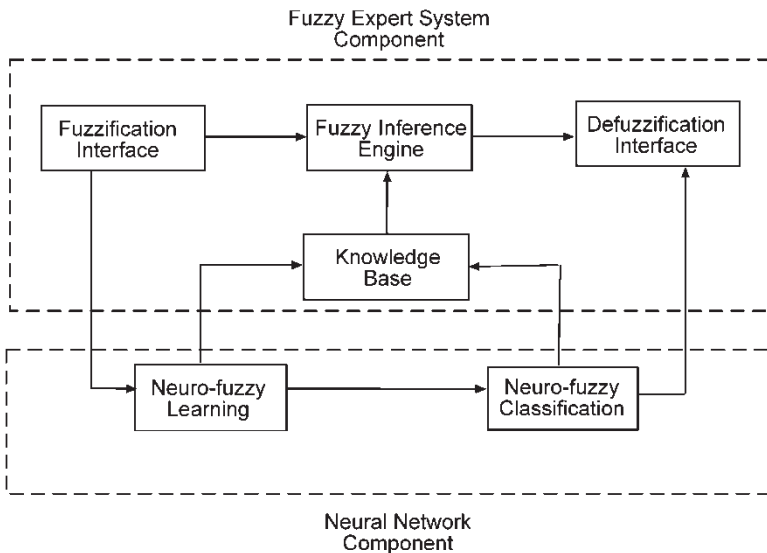


Figure 3. The architecture of the hybrid neuro-fuzzy image classification system.

3. neuro-fuzzy learning, and
4. neuro-fuzzy classification and defuzzification.

2.1. Fuzzification interface

Inputs into a neural network system are usually not fuzzy numbers. Therefore, the first step in building the neuro-fuzzy system is to fuzzify the input remote sensing data into a set of fuzzy numbers. The fuzzification of a single value is usually accomplished by applying certain fuzzy membership functions. Because a single sigmoid function does not adequately represent a fuzzy number with a closed interval and a triangle function does not insure that an input value will be covered by at least one fuzzy set, the Gaussian function is employed in the improved system, as was the case in the FIN based neuro-fuzzy system.

$$\mu A_{ij} = \exp\left(-1/2 \frac{(c_{ij} - x_j)^2}{\sigma_{ij}^2}\right) \tag{3}$$

where μ is the membership grade for fuzzy number A_{ij} , c_{ij} is the centre parameter of the Gaussian function relating i th target information class with j th input variable x_j (i.e. the centre of one data cluster for a SOM), and σ_{ij} represents the corresponding standard deviation parameter (i.e. the fuzzy set boundary of the cluster). For an input pixel, the new system uses the fuzzy membership grade, rather than the Euclidean distance, to determine the closeness of the input pixel values to the corresponding cluster. This enables increased accuracy in the assignment of a pixel to clusters with different dispersions as shown in figure4, where P is a dispersed cluster and Q is a compact cluster. Using the minimum Euclidean method, a pixel i of cluster P is classified as a member of cluster Q because it is ‘closer’ to the centre of cluster Q based on Euclidean distance. Conversely, this pixel is assigned to cluster P using the fuzzy membership function, due to the consideration of data distribution of the cluster.

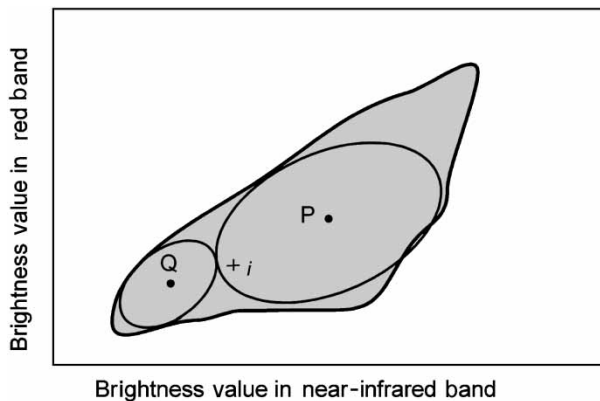


Figure 4. Comparison of pixel assignment strategy between Kohonen’s self-organizing map (SOM) and improved fuzzy neuro-fuzzy system using a hypothetical feature space plot. P is the centre of a dispersed cluster and Q is that of a compact cluster. Using the minimum Euclidean method, a pixel i of cluster P would be incorrectly classified as a member of cluster Q based on Euclidean distance. The pixel will be assigned to cluster P if using the fuzzy membership function.

2.2. Fuzzy inference engine and knowledge base

A Gaussian membership function allows the determination of the fuzzy membership grade based on one spectral band. To determine whether a pixel should be assigned to a particular land use or land cover class, it is necessary to consider the brightness values of all the bands associated with the pixel. Instead of using a single fuzzy set, the improved neuro-fuzzy system uses a fuzzy 'if-then' rule to decide the overall membership grade of the pixel. A fuzzy 'if-then' rule is actually a combination of fuzzy sets defined for all the bands and connected by a specific type of fuzzy logic operator Ω :

If band 1 is fuzzy number A_{i1} Ω band₂ is fuzzy number A_{i2} Ω ... band_j is fuzzy number A_{ij} ... Ω band n is a fuzzy number A_{in} , then the pixel y is class C_i ,

where n is the total number of input bands, C_i is the i th target information class, and i and j are the same as defined before. The fuzzy logic operator Ω plays an important role in determining the overall membership grade for the pixel and the degree of the rule being activated (Nomura and Miyoshi 1995, Benitez *et al.* 1997). Because an *interactive-or* fuzzy connective operator appears semantically complicated, and the algebraic product fuzzy intersection operator tends to underestimate the overall membership grade of a fuzzy rule, the standard fuzzy intersection (i.e. the minimum function) seems to be an appropriate connective operator. However, the standard fuzzy intersection assigns the lowest membership grade of all the bands to the pixel as the overall membership grade, whereas the standard union operator utilizes only the band with the highest degree of membership. Both use the extreme values but do not take the balance of all the bands into consideration. Therefore, a new fuzzy *and-or* operator in the form of a geometric mean was used when deriving the overall membership grade:

$$\text{and-or } (a_1, a_2, \dots, a_n) = (a_1 \cdot a_2 \dots a_n)^{1/n} \quad (5)$$

where a is the membership grade for an input band, and n is the number of bands. By geometrically averaging the membership grades of all the bands, the overall membership grade α_i of a fuzzy rule concerning i th target information class becomes:

$$\alpha_i = \left(\sum_{j=1}^n \exp \left(-1/2 \frac{(c_{ij} - x_j)^2}{\sigma_{ij}^2} \right) \right)^{1/n} \quad (6)$$

where i , j , n , c and σ are the same as previously defined. The *and-or* operator is a fuzzy operator sitting between fuzzy *and* (intersection) and fuzzy *or* (union) operators and enables a low membership grade of one band to be compensated by a high membership grade of another band, so that a missing or noisy value in one band will not heavily affect the classification output of the entire pixel. In addition, the *and-or* operator is an 'idempotent' function, such that *and-or* (a, a, a) = a . For example, suppose the membership grade for all three bands is 0.5, the geometric mean of all the bands will produce an overall membership grade of 0.5, a more logical result than that derived from the algebraic product used in the Nomura and Miyoshi's (1995) model (i.e. 0.125).

If the number of neurons in competitive layers is greater than that of neurons in the output layers, multiple rules can be extracted for one output class. For example, for a network having twice as many competitive neurons as output neurons, two fuzzy 'if-then' rules consisting of fuzzy sets connected by *and-or* operators can be

extracted and joined with a fuzzy union (*or*) operator. By applying the standard fuzzy union operation (i.e. the maximum function) to the two rules joined by the *or* operator, the output of the neuron representing the final membership grade of a class in the output layer can be derived. The rules derived for all the information classes construct the knowledge/rule base, which can be used to perform land use/land cover classification of the image.

2.3. Neuro-fuzzy learning

The topological structure of the underlying neural network predetermines the number and the form of the fuzzy 'if-then' rules being used in the fuzzy inference engine. Each rule consists of several fuzzy sets (numbers) specified by Gaussian functions. For the system to yield an accurate image classification, the two parameters c and σ of each Gaussian function are fine-tuned by the fuzzy neural network learning algorithm. The updating of the c parameter is accomplished by the traditional LVQ learning algorithm (Kohonen *et al.* 1996):

$$\Delta c_{ij} = \eta(x_j - c_{ij}), \text{ if } x \text{ and } c \text{ belong to same class} \quad (7)$$

$$\Delta c_{ij} = -\eta(x_j - c_{ij}), \text{ otherwise}$$

where η is the learning rate, a small positive constant that decreases with training time, and i, j, x and c are the same as defined above. By intentionally using the true target class information, the supervised LVQ learning method has the effect of moving c towards the centre of the correctly classified data cluster and away from the centre of incorrectly classified data clusters. Compared to an unsupervised self-organizing algorithm, this mechanism increases the speed of the learning process and improves classification accuracy.

A new learning rule was proposed to adjust the weight σ that simulates the standard deviation (fuzzy set boundary) parameter of the Gaussian function:

$$\Delta \sigma_{ij} = \eta(|c_{ij} - x_j| - \sigma_{ij}), \text{ if } x \text{ and } c \text{ belong to same class} \quad (8)$$

$$\Delta \sigma_{ij} = 0, \text{ otherwise}$$

where η is the learning rate, and i, j, x and c are defined as before. The new update rule solves the problems encountered by Nomura and Miyoshi's (1995) learning rule for the standard deviation parameter σ . By dropping the square operators in the original learning rule, this method updates the standard deviation parameter directly rather than through the adjustment of the variance parameter. The replacement of the squared difference with the absolute value of the difference between an input value x and the corresponding centre parameter c allows training to take place using raw remote sensing data in the range of 0 to 255 without crashing the system. Even with scaled data in the range of 0 to 1, this method reached a system convergence quickly due to the absence of squaring operations that diminish the weight adjustment.

2.4. Neuro-fuzzy classification and defuzzification

With all the fuzzy 'if-then' rules formed and their parameters automatically fine-tuned by the fuzzified neural network, the neuro-fuzzy system can finally classify pixels of remote sensing images into land use/land cover classes. In a traditional image processing system, the result is a single classification map where a pixel can be assigned to only one land use/land cover type (Jensen 1996). However, the

output of the neuro-fuzzy classification is a fuzzy membership grade. The resulting fuzzy membership maps provide more information than a traditional classification map. If properly utilized, they will provide information about various constituent classes found in a mixed pixel (Jensen 1996).

Another attraction of the neuro-fuzzy classification is that the classification of the image can be carried out without using the trained neuro-fuzzy system. Once the fuzzy 'if-then' rules are extracted and stored, the neuro-fuzzy system is no longer needed. As an alternative, classification can easily be accomplished by directly applying the derived fuzzy 'if-then' rules to the remote sensing images to be classified.

Defuzzification is the reverse of the fuzzification process. It can be used to convert the fuzzy membership maps into a hard classification map. Defuzzification is obtained by comparing all the membership grades of a pixel and then assigning the pixel to the class with the maximum membership grade. The weighted linear combination defuzzification method is not used as in other neuro-fuzzy systems. The output of a weighted linear combination method is a continuous numerical (ratio) value. Results of an image classification problem are discrete nominal categories.

3. Analysis results and discussion

To test its ability to solve real world classification problems, the improved neuro-fuzzy image classification system developed in this study was employed to classify a high spatial resolution remotely sensed image. The classification accuracy was quantitatively assessed and compared with the results obtained from statistical and neural network approaches. Based on the interpretation of the neuron weights, fuzzy 'if-then' rules were generated to obtain an understanding of the image classification process involved in the neural network component. The system also provided a capability to fine-tune the parameters for each of the fuzzy system components.

3.1. Study area and dataset

The study area was Jacksonville Beach, FL, USA. The intra-coastal waterway passes through the centre of the study area and flows to the Atlantic Ocean. There is a gradual change of land cover from water to wetlands, to upland forests, to urban infrastructure. The land cover transition zones complicate the image classification task. A National Aerial Photography Program (NAPP) colour-infrared digital orthophotoquad image (figure 5) of the study area with $1\text{ m} \times 1\text{ m}$ spatial resolution was used to perform the land use/land cover classification. The digital image consists of three bands of digitized data: red, green, and near-infrared. For all practical purposes, the digital NAPP image is similar to other high spatial resolution satellite remote sensor data.

3.2. Neuro-fuzzy training and fuzzy rule extraction

Due to the limited amount of spectral information in the NAPP digital image (i.e. only three bands), the land cover classes to be identified included only water, unforested wetland, upland forest, and urban/built-up land. After examining the histograms of the brightness values, it was determined that two rules would be generated for each land cover type. This decision was made to model the bimodal distribution patterns in the image. Therefore, the topological structure of the system was configured as 3:8:4 (i.e. 3 input, 8 competitive and 4 output neurons) as

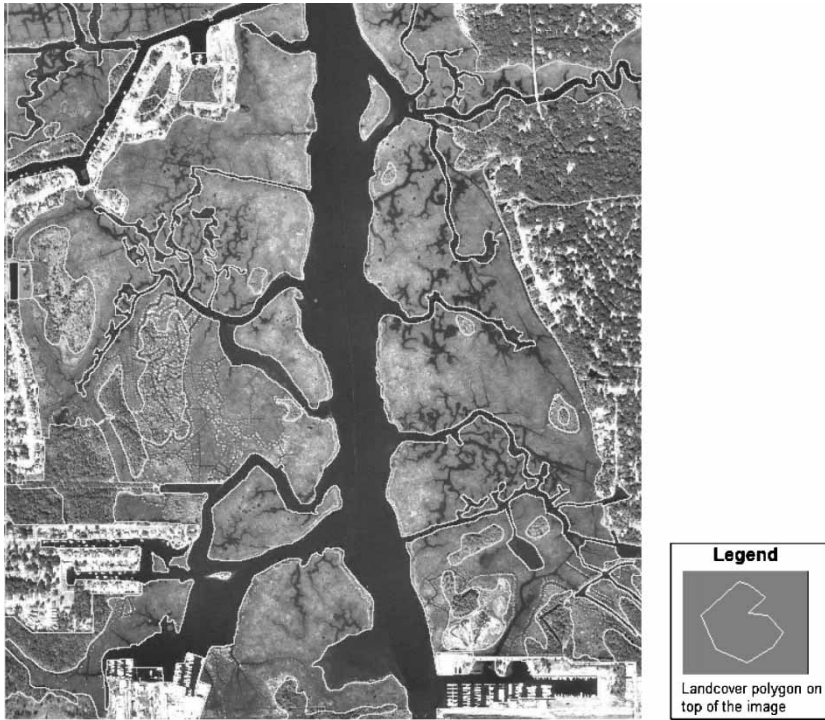


Figure 5. Land cover polygons derived from visual interpretation and digitization based on the underlying National Aerial Photograph Program (NAPP) image at the Jacksonville Beach, FL, USA study area.

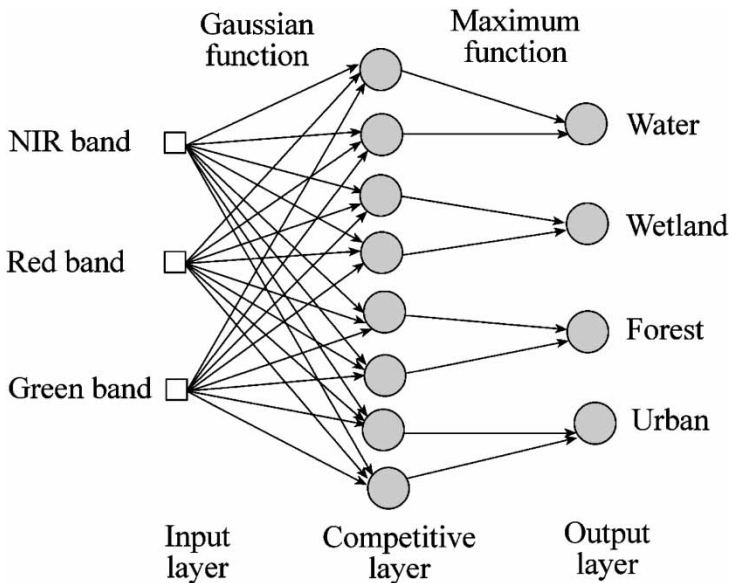


Figure 6. The 3:8:4 topological structure of the fuzzy neuro-system used in this study.

shown in figure 6. The output and competitive layers are not fully connected as shown in the illustration.

For the neuro-fuzzy system to learn the image classification rules, 47 training sites with ground truth representing homogeneous examples of known land cover types were selected. The three-band brightness values and the associated information classes of the training data were then provided to the system as input-output pairs to perform supervised training. The system was able to achieve convergence without crashing through iterative learning.

An understanding of the knowledge used by the neuro-fuzzy system to perform the image classification was obtained by interpreting the weights automatically extracted from the trained system as shown in table 1. There are two sets of fuzzy parameters represented by the weights. These two sets jointly determine the Gaussian membership functions of the fuzzy numbers depicting the data cluster. The fuzzy numbers provide not only the spectral centroid/mean information of the data clusters (table 1(a)), but also their fuzzy set boundary (embodied by the standard deviation parameters in table 1(b)) in the associated feature space.

Based on the interpretation of these fuzzy numbers, the knowledge used by the neuro-fuzzy classification system can be derived in the form of fuzzy ‘if-then’ rules. Each fuzzy ‘if-then’ rule is composed of the fuzzy numbers of different bands connected by the *and-or* logic operator. Each column in table 1(a) and (b) jointly determined one fuzzy ‘if-then’ rule of a class. Equation (6) determines that the same column *i* of the two tables has to be used to establish a fuzzy rule for class C_i . This rule depicts the spectral characteristics of a data cluster of that class. For example, the first column of the two tables establishes a fuzzy ‘if-then’ rule for the first data cluster of the water land cover type. For each land cover type, two fuzzy ‘if-then’ rules are derived and joined by an *or* fuzzy logic connective operator. The fuzzy ‘if-then’ rules generated for the water land cover type were:

If the NIR band is a fuzzy number 19.070 (with a fuzzy set boundary of 0.307) *and-or* the red band is a fuzzy number 43.256 (with a fuzzy set boundary of 0.653) *and-or* the green band is a fuzzy number 53.858 (with a fuzzy set boundary of 0.466), then the pixel is a water class *or*

Table 1. Fuzzy rule parameters extracted from the hybrid neuro-fuzzy system.
(a) The extracted centre parameters for the fuzzy numbers in the fuzzy rules.

	Water		Wetland		Forest		Urban	
	c1	c2	c1	c2	c1	c2	c1	c2
NIR	19.070	20.254	52.234	83.861	84.974	148.230	177.597	223.765
Red	43.256	46.053	74.065	102.858	59.331	87.329	195.683	223.991
Green	53.858	53.718	104.747	146.575	75.340	126.001	217.097	226.237

(b) The extracted fuzzy set boundary parameters for the fuzzy numbers in the fuzzy rules.

	Water		Wetland		Forest		Urban	
	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
NIR	0.307	1.789	6.800	5.585	10.030	16.819	9.427	7.415
Red	0.653	4.941	7.845	4.732	4.346	11.532	10.097	5.737
Green	0.466	6.894	12.021	4.534	11.557	12.309	5.909	4.029

If the NIR band is a fuzzy number 20.254 (with a fuzzy set boundary of 1.789) *and-or* the red band is a fuzzy number 46.053 (with a fuzzy set boundary of 4.941) *and-or* the green band is a fuzzy number 53.718 (with a fuzzy set boundary of 6.894), then the pixel is a member of water class.

Similarly, fuzzy rules for wetland, forest, and urban land cover were derived using the parameters of the corresponding columns in table 1(a) and (b). These fuzzy rules form the knowledge base and offer an understanding of how the neuro-fuzzy system performs the image classification process. For example, the fuzzy ‘if-then’ rules demonstrate that pixels with low brightness values in all three bands are usually classified as water, while pixels with high brightness values in all three bands are classified as urban/built-up lands. Forests generally exhibit a high

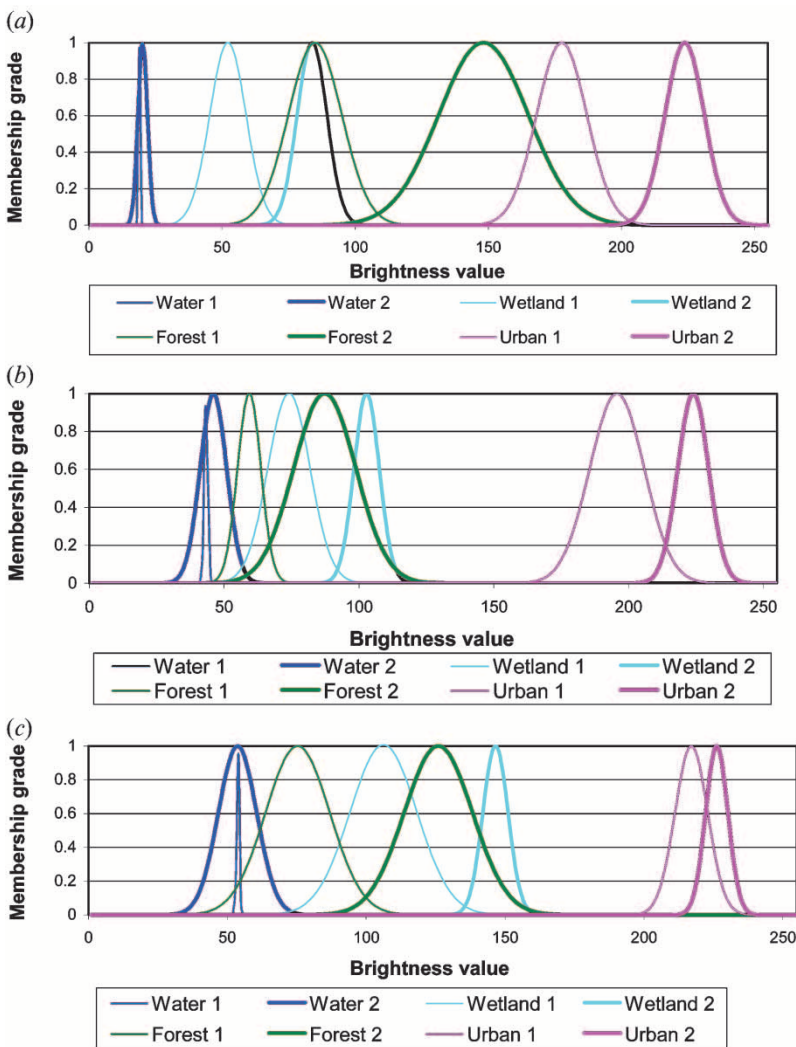


Figure 7. Fuzzy membership graphs obtained from the extracted fuzzy ‘if-then’ rules. (a) Near-infrared band, (b) red band, (c) green band.

reflectance in the near-infrared band but have low reflectance in the red and the green bands, as expected (Jensen 2000).

These fuzzy 'if-then' rules can also be visualized as fuzzy membership graphs of different bands (figure 7). The visualization of the fuzzy 'if-then' rules reveals that all the land cover types except water demonstrate patterns that can be better modelled by two rules than by one rule. The water class can be effectively modelled by just one rule because it is relatively homogenous and exhibits a single mode in its data distribution. In fact, the first rule for the water class (water 1) is actually encompassed by the second rule for the water class (water 2) because all the Gaussian functions for water 1 have very small fuzzy boundaries (<0.7) and similar centre parameters with those of water 2 for all three bands (figure 7(a)–(c)). It was found that modifying the derived knowledge by eliminating the redundant rule (such as the first rule for the water class) did not affect the final classification accuracy. It was also possible to combine the two rules for the urban class because they were well separated from the other classes as shown in all three fuzzy membership graphs (figure 7(a)–(c)). In addition, the employment of more than one rule helps differentiate land cover classes that are difficult to separate. The unsupervised self-organizing approach incorporated in the fuzzy LVQ learning algorithm attempts to arrange the multiple data clusters in the feature space with the best separation. For instance, although there is total overlap between wetland 2 and forest 1 in the near-infrared band (figure 7(a)), their Gaussian membership functions in the red and the green bands (figure 7(b), (c)) make it possible to isolate these two classes. Complex datasets may require possibly three or more rules to be used for each land cover type.

Remote sensing classification may be performed using the trained neuro-fuzzy system as an ordinary neural network. Classification may also be accomplished by simply employing the derived and simplified fuzzy 'if-then' rules in the knowledge base. Each fuzzy 'if-then' rule of a class can be applied to the image using equation (6) with the associated parameters to calculate the overall membership grades for each class. The final membership grade for each class is determined by the standard fuzzy union (the maximum function). The defuzzification process then assigns the pixel to the class with the highest overall membership grade. The advantage of using the fuzzy expert system rather than the trained neural network is the possible improvement of the system through the incorporation of human expertise.

3.3. Accuracy assessment

A back-propagation based neural network and a maximum likelihood statistical classification algorithm were also used to classify the same digital NAPP image for comparison purpose. The back-propagation based neural network is the same as that used by Benitez *et al.* (1997). We did not test the classification accuracies of Vuorimaa's (1994) FSOM based system and Nomura and Miyoshi's (1995) FIN based system, because neither system was able to achieve convergence even during image training stage. The classification accuracies obtained were compared to those derived using the improved neuro-fuzzy classification algorithm. A total of 400 test points (256 random sample points and 144 stratified random sample points) with ground truth were selected for accuracy assessment based on binomial probability theory.

Table 2 displays overall accuracy and the Kappa coefficients (K_{hat} statistics) of

Table 2. Image classification accuracy for the three methods.

	Neuro-fuzzy system	Back-propagation	Maximum likelihood
Overall accuracy	90.75%	83.75%	81.25%
Kappa (K_{hat} statistic)	0.8695	0.7832	0.7285
Kappa variance	0.000411	0.000605	0.000771

the three classification results based on the test points. The overall accuracy for the improved neuro-fuzzy image classification system was 90.75%. The overall classification accuracy using the back-propagation based neural network was 83.75%. The overall classification accuracy using the maximum likelihood classification was 81.25%.

K_{hat} is believed to be a better representation of the general quality of classification because it removes the effects caused by the differences in sample size and also accounts for the off-diagonal elements in the error matrix (Rosenfield and Fitzpartrick-Lins 1986). The K_{hat} statistic for the improved neuro-fuzzy based classification was 0.8695. The K_{hat} statistic for the back-propagation based neural network classification was 0.7832. The K_{hat} statistic for the maximum likelihood classification was 0.7285. According to Fleiss (1981), Kappa coefficients >0.75 suggest strong agreement above chance. Landis and Koch (1977) suggest that Kappa coefficients >0.81 are almost perfect. To statistically compare the Kappa coefficients, a significance Z-test was performed between every two independent K_{hat} statistics (Rosenfield and Fitzpartrick-Lins 1986). Table 3 lists the Z-scores in a pairwise manner between the three classification methods. The results indicate that the neuro-fuzzy based classification was significantly better than both the back-propagation based classification and the maximum likelihood classification at the 99% confidence level.

4. Conclusions and future study

Neural networks and fuzzy systems exhibit many advantages and provide alternative methods for processing remote sensing data. However, neither neural networks nor fuzzy expert systems have been widely applied by the remote sensing community because of their liabilities. The 'black box' problem of neural networks and the knowledge automations problem associated with fuzzy systems have impeded the widespread adoption of these two methods being applied to the classification of complex remote sensing data.

This research investigated the possibility of solving these two problems through the development of an improved neuro-fuzzy image classification system based on the synergism between neural networks and fuzzy system technologies. The advantages and limitations of three existing systems were taken into account in

Table 3. Results of a Kappa Z-test for the error matrices from the three approaches.

	Hybrid neuro-fuzzy	Back-propagation	Maximum likelihood
Hybrid neuro-fuzzy		2.707 (SS)	4.101 (SS)
Back-propagation			1.474 (NS)

NS – Difference not significant at 95% confidence level ($Z < 1.960$), S – Difference is significant at 95% confidence level ($Z > 1.960$), SS – Difference is significant at 99% confidence level ($Z > 2.575$).

order to develop an operational neuro-fuzzy image classification system. A newly updated rule to derive fuzzy set boundary parameters for the Gaussian membership functions was proposed and demonstrated to be superior to the original method proposed by Nomura and Miyoshi (1995) in terms of stability and learning speed. Using an *and-or* fuzzy logic operation in the form of a geometric mean, a more comprehensible and realistic fuzzy membership grade was obtained.

The neuro-fuzzy system based on the fuzzified LVQ algorithm, which combines the training strategies of both supervised learning and unsupervised learning, was able to significantly outperform traditional statistical (maximum likelihood) and back-propagation based neural network approaches. The improved neuro-fuzzy system was also able to help establish comprehensible fuzzy 'if-then' rules and automate the associated fuzzy set parameters, so that the decision processing of neural network learning is made transparent. The associated fuzzy membership graphs offer alternative means to explore the remote sensing data being classified. The graphs not only provide the analyst with the centroid information of the data cluster, but also their fuzzy set boundaries as well. The automation of the associated fuzzy set parameters for the derived fuzzy 'if-then' rules alleviates to some extent the knowledge acquisition problems of the fuzzy expert system. After fuzzy 'if-then' rules have been derived, use of a neural network is no longer necessary. The derived fuzzy rules formed a knowledge/rule base, which can be employed directly to perform image classification in the experimental area and other places with similar conditions. Human expertise can also be incorporated into the classification process to improve system performance. The integration of a neural network learning algorithm with the fuzzy system symbolic representation has opened the black box of the neural network and simultaneously facilitates the knowledge acquisition automation of the fuzzy expert system.

Though the associated fuzzy set parameters can be automatically derived using the improved neuro-fuzzy system, at present, the form and the number of rules to be extracted are still manually predetermined. With the incorporation of human expertise, redundant rules can be pruned out based on the analysis of their associated fuzzy set parameters. For a complete breakdown of the knowledge acquisition bottleneck of fuzzy systems, the automatic determination of the form and the number of fuzzy rules to be extracted still needs to be addressed. This constitutes the future research of this study.

Acknowledgment

The authors would like to extend our appreciation to Ms. Annie Hsu and the reviewers of this manuscript. Their valuable comments were very constructive in improving the quality of the paper.

References

- ABE, S., and LAN, M., 1996, Effective method for fuzzy rule extraction from numerical data. In *Fuzzy Logic and Neural Network Handbook*, edited by C. H. Chen (New York: McGraw-Hill), pp. 7.1–7.33.
- ANDREWS, R., DIEDERICH, J., and TICKLE, A. B., 1995, Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge-based Systems*, **8**, 373–389.
- BENEDIKTSSON, J. A., and SVEINSSON, J. R., 1997, Feature extraction for multi-source data classification with artificial neural networks. *International Journal of Remote Sensing*, **18**, 727–740.

- BENITEZ, J. M., CASTRO, J. L., and REQUENA, I., 1997, Are artificial neural networks black boxes? *IEEE Transactions on Neural Networks*, **8**, 1156–1164.
- BINAGHI, E., MADELLA, P., MONTESANO, M. G., and RAMPINI, A., 1997, Fuzzy contextual classification of multi-source remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, **35**, 326–339.
- CIVCO, D. L., 1993, Artificial neural network for land cover classification and mapping. *International Journal of Geographical Information System*, **7**, 173–186.
- FAUSETT, L., 1994, *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications* (Englewood Cliffs, NJ: Prentice Hall).
- FLEISS, J. L., 1981, *Statistical Methods for Rates and Proportions*, 2nd edn (New York: John Wiley & Sons).
- FOODY, G. M., MCCULLOCH, M. B., and YATES, W. B., 1995, Classification of remotely sensed data by an artificial neural network: issues related to training data characteristics. *Photogrammetric Engineering and Remote Sensing*, **61**, 391–401.
- GONG, P., 1996, Integrated analysis of spatial data from multiple sources: using evidential reasoning and artificial neural network techniques for geological mapping. *Photogrammetric Engineering and Remote Sensing*, **62**, 513–523.
- HAGAN, M. T., DEMUTH, H. B., and BEALE, M., 1996, *Neural Network Design* (Boston, MA: PWS).
- HAYKIN, S., 1994, *Neural Networks: A comprehensive foundation* (New York: Macmillan).
- HEERMANN, P. D., and KHAZENIE, N., 1992, Classification of multispectral remote sensing data using a back-propagation neural network. *IEEE Transactions on Geosciences and Remote Sensing*, **30**, 81–88.
- HEWITSON, B. C., and CRANE, R. G., 1994, Looks and use. In *Neural Nets: Applications in geography*, edited by B. C. Hewitson and R. G. Crane (Dordrecht/Boston/London: Kluwer Academic), pp. 1–9.
- HORNIK, K., STINGHCOMBE, M., and WHITEE, H., 1989, Multilayer feedforward networks are universal approximators. *Neural Networks*, **2**, 359–366.
- JENSEN, J. R., 1996, *Introductory Digital Image Processing: A remote sensing perspective* (Upper Saddle River, NJ: Prentice-Hall).
- JENSEN, J. R., 2000, *Remote Sensing of the Environment: An Earth resource perspective* (Upper Saddle River, NJ: Prentice-Hall).
- JENSEN, J. R., and COWEN, D. C., 1999, Remote sensing of urban/suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing*, **65**, 611–622.
- JENSEN, J. R., QIU, F., and JI, M., 1999, Predictive modeling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensing data. *International Journal of Remote Sensing*, **20**, 2805–2822.
- KLIR, G. J., and YUAN, B., 1995, *Fuzzy Sets and Fuzzy Logic: Theory and application* (Upper Saddle River, NJ: Prentice Hall).
- KOHONEN, T., 1989, *Self-Organization and Associate Memory*, 3rd edn (New York: Springer-Verlag).
- KOHONEN, T., HYNINEN, J., KANGAS, J., LAAKSONEN, J., and TORKKOLA, K., 1996, LVQ_PAK: *The Learning Vector Quantization Program Package*. Technical report of Helsinki University of Technology, Finland, 26 p.
- KULKARNI, A. D., and LULLA, K., 1999, Fuzzy neural network for supervised classification: multispectral image analysis. *Geocarto International*, **14**, 41–49.
- LANDIS, J., and KOCH, G. G., 1977, The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.
- LEIN, J. K., 1997, *Environment Decision Making: An information technology approach*, Malden (Malden, MA: Blackwell Science).
- LIPPMANN, R. P., 1987, An introduction to computing with neural nets. *IEEE ASSP Magazine*, **4**, 4–22.
- LLOYD, R., 1996, *Spatial Cognition, Geographic Environments* (Dordrecht: Kluwer Academic).
- MEDSKER, L. R., 1994, *Hybrid Neural Network and Expert Systems* (Boston, MA: Kluwer Academic).
- NAUCK, D., KLAWOON, F., and KRUSE, R., 1997, *Foundation of Neuro-Fuzzy Systems* (Chichester: John Wiley & Sons).
- NOMURA, T., and MIYOSHI, T., 1995, An adaptive rule extraction with the fuzzy

- self-organizing map: a comparison with other methods. *Proceedings of the Third International Symposium on Uncertainty Modeling and Analysis and Annual Conference of the North American Fuzzy Information Processing Society, IEEE Computer Society, September 17–20, 1995* (College Park, MD: IEEE), pp. 311–316.
- OPENSHAW, S., and OPENSHAW, C., 1997, *Artificial Intelligence in Geography* (Chichester: John Wiley & Sons).
- ROLI, F., SERPICO, S. B., and VERNAZZA, G., 1996, Neural networks for classification of remotely sensed images. In *Fuzzy Logic and Neural Network Handbook*, edited by C. H. Chen (New York: McGraw-Hill), pp. 15.1–15.28.
- ROSENFELD, G. H., and FITZPATRICK-LINS, K., 1986, A coefficient as measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, **52**, 223–227.
- VUORIMAA, P., 1994, Fuzzy self-organizing map. *Fuzzy Sets and Systems*, **66**, 223–231.
- WANG, F., 1990, Improving remote sensing image analysis through fuzzy information representation. *Photogrammetric Engineering and Remote Sensing*, **56**, 1163–1169.
- ZADEH, L. A., 1965, Fuzzy sets. *Information and Control*, **8**, 338–353.