Integrated Analysis of Spatial Data from Multiple Sources: Using Evidential Reasoning and Artificial Neural Network Techniques for Geological Mapping

P. Gong

Abstract

As the availability of digital spatial data, other than from remote sensing, increases, it becomes increasingly important to develop algorithms to handle both remote sensing and other spatial data. For classification purposes, commonly used remote sensing algorithms such as the maximum-likelihood classifier and the minimum-distance classifier can only be used to deal with spatial data of interval and ratio scales. They are not applicable to spatial data of nominal or ordinal scale as exemplified by data digitized from a categorical map. Bayesian theory, mathematical theory of evidence, and artificial neural networks, on the other hand, are capable of handling data with any measurement scale. In this paper, we introduce an evidential reasoning and a back-propagation feed-forward neural network algorithm and evaluate their applications to classification problems. A multisource data set including Landsat Thematic Mapper, aeromagnetic, radiometric, and gravity data has been used in the classification of four rock types in Melville Peninsula, Northwest Territories, Canada. The highest overall accuracy of 96.0 percent and average accuracy of 92.1 percent were achieved with the neural network algorithm while the evidential reasoning method produced an overall accuracy of 94.7 percent and average accuracy of 89.3 percent. The evidential reasoning method resulted in three highest individual class accuracies out of the four classes.

Introduction

As the amount of digital data increases, it is desirable to use computers to analyze simultaneously remotely sensed data of different resolutions and other spatial data, including digital maps at different measurement scales. Spatial data from different sources have different accuracies and measurement scales. Their accuracies are dependent on methods of data collection, manipulation, interpretation, and presentation. For example, the accuracy of a map depends not only on its contents and map scale, but also on how the original map data were acquired, processed, and converted into digital forms. The accuracies of remotely sensed data depend largely on their spatial and radiometric resolutions. The measurement scale of spatial data varies from nominal and ordinal to interval and ratio (Robinson et al., 1984). For example, maps showing land-cover types or geological structures present data at a nominal scale while a slope-class map shows different slopes with an ordinal scale such as flat, middle, and steep. Remote sensing images, on the other hand, record radiance of surface targets with a ratio scale. From a user's point of view, it is difficult to use all these digital data together because they have different levels of reliability, uncertainty, and completeness.

Developing algorithms that are capable of handling both remote sensing and other spatial data has been an active research area, referred to as multisource data integration (NCGIA, 1989). A number of data integration strategies have been proposed (e.g., Corr et al., 1989; Cibula and Nyquist, 1987; Marble and Peuquet, 1983; Hutchinson, 1982). However, none of these strategies can handle data at all measurement scales. Recent works have demonstrated the value of mathematical theory of evidence for land-cover classification using data at many measurement scales (Lee et al., 1987; Chung and Moon, 1991; Peddle, 1993; Wang and Civco, 1994). As demonstrated in land-cover classification of remote sensing and topographic data such as slope and aspect (Benedikttson et al., 1993), neural network techniques can handle data at any measurement scale although data preprocessing such as normalization sometimes needs to be done. Peddle et al. (1994) compared evidential reasoning with two neural network algorithms for land cover classification of SPOT multispectral data and topographic variables. They reported that neural network algorithms produced better overall classification accuracies than did the evidential reasoning method. Difficulties were encountered, however, in classifying ecological land systems when a neural network algorithm was applied to digital map data such as forest cover maps, digital elevation data, slope and aspect, and soil data (Chen et al., 1993; Gong et al., 1994).

In this paper, we first review some commonly used classification methods as tools for integrated analysis of multisource spatial data. Their limitations are briefly examined. We then introduce the evidential reasoning and neural network methods used in this research. Finally, we concentrate on the application and assessment of the two methods for geological mapping.

Department of Environmental Science, Policy, and Management, Center for Assessment and Monitoring of Forest and Environmental Resources, 145 Mulford Hall, University of California, Berkeley, CA 94720.

Classification: A Technique for Environmental Assessment and Natural Resources Management

Classification is a process of abstraction and generalization from data collected about certain phenomena to improve our understanding. It is an important component of integrated analysis of multisource spatial data (Gong, 1994). It involves grouping individual measurements and labeling each group as a class according to certain rules of similarity in the measurement space. For spatial data collected about our environment and natural resources, classification serves for reducing the level of complexity in spatial data and making it easier for environmental planners, resources managers, and decision-makers to understand, use, and communicate about such data. To achieve this goal, we first qualitatively observe and quantitatively measure the spatial phenomena with available instruments. As a result, we obtain descriptions and measurements about every individual target of interest. These descriptions and measurements constitute a set of discriminant variables, often called features or pieces of evidence.

The question is, given a set of features or pieces of evidence, denoted by $\mathbf{X} = \{x_1, x_2, ..., x_m\}$, observed or measured with different methods from the object space, how can a computer be used to decide which class, among a set of k classes $\{C_1, C_2, ..., C_k\}$, is the most appropriate one to which $\mathbf{X}$ should be classified. More specifically, two questions can be asked:

- to which particular class does a target of interest belong? That is, to classify each target into only one class; and
- to what extent does a target belong to a class? That is, to allow partial membership in each class.

The second type of decision is more quantitative and in many circumstances more realistic than the first type. Answering the first question, however, is what we traditionally do in classification. Knowing the answer to the second question, we can easily solve the first question by comparing the partial membership values. The second type of decision making is a more recent subject that involves the determination and use of partial memberships to help determine class uncertainties and ambiguities in class definitions (e.g., Zadeh, 1965; Shafer, 1976; Bezdek *et al.*, 1984; Liu and Burrough, 1987).

To help answer these two questions, some classification concepts used in remote sensing are useful. Image classification algorithms can be divided into two groups based either on whether a training process is needed or on whether a parametric model is used (Swain and Davis, 1978). If training is required, the classification is a supervised one. Otherwise, it is unsupervised. If a parametric model is required, the algorithm is called a parametric one. Otherwise, it is non-parametric. We therefore can categorize classification algorithms into four groups:

- supervised parametric classifier,
- unsupervised parametric classifier,
- supervised non-parametric classifier, and
- unsupervised non-parametric classifier.

For instance, the commonly used maximum-likelihood classifier (MLC) is a supervised parametric classifier; the k-means clustering and the ISODATA clustering algorithms are unsupervised parametric classifiers (Richards, 1986); the non-parametric Bayesian algorithm is a supervised non-parametric classifier (Gong and Dunlop, 1991; Skidmore and Turner, 1988); and a histogram-based clustering is an unsupervised non-parametric classifier (Richards, 1986).

While each type of classifier has its own advantages, their use in integrated analysis of multisource spatial data is limited by one or more factors, including class probability distribution, form of knowledge and procedure for classification, algorithm complexity, and data characteristics such as measurement scale, resolution, and dimensionality. For example, the MLC assumes that the class probability distribution is multivariate normal. Many clustering algorithms and the MLC can be applied only to data at ratio and interval measurement scales. Singularities of covariance matrices stemming from the use of multiple resolution data restrict the use of the MLC. Non-parametric classifiers may allow only a small number of features or a small number of samples to be classified due to the limitations of computer memory or computation. Some algorithms require a *priori* class probabilities as a form of classification knowledge known before classification. Such knowledge, however, is often difficult to obtain. Therefore, alternative algorithms must be developed to circumvent these problems.

Classification Based on Evidential Theory

Let a vector $\mathbf{X} = (x_1, x_2, ..., x_n)$ denote a set of observations or measurements made at a particular location. $\mathbf{X}$ is a set of features or n pieces of evidence. Classification can be considered as a multivalued mapping, $\Gamma: E \rightarrow 2^C$, that associates each element $\mathbf{X}$ in E with a set of elements in 2^C. E is the feature space, also called observation space or evidence space, $C = \{C_1, C_2, ..., C_k\}$ is the class space whose elements are mutually exclusive, and 2^C is the *universe of discourse* or the *frame of discernment*, i.e., the set that contains all possible sets consisting of elements in C and the empty set Φ.

To realize the mapping: $\Gamma: E \rightarrow 2^C$, in classification it is often simplified to: $\Gamma: E \rightarrow C$. We use C instead of 2^C because our interest is focused on the individual elements in C, i.e., the case of singleton hypotheses. For example, our purpose usually is to find the probabilities of each individual class in $C = \{$Urban, Agriculture, Forest, Water$\}$: $P(U)$, $P(A)$, $P(F)$, and $P(W)$, but we are not interested in knowing $P(\{U, A\})$, $P(\{A, F, W\})$, ..., etc.

In evidential theory (Shafer, 1976), a *basic probability assignment* (BPA) of C, denoted by m: $C \rightarrow [0, 1]$, is defined as

$$m(A) = \sum_{i: f(x_i)=A} p(x_i) \qquad (1)$$

where f is the mapping function from a subspace of E to C, A is a subset of C which is called *a focal element*, and $p(x_i)$ is the probability density of x_i in a subspace of E. The "BPA" is also referred to as a mass function to distinguish it from the probability distribution.

A mass function has the following property:

$$\sum_{A \subset C} m(A) = 1 \qquad m(\Phi) = 0.$$

The probability distribution of C can be estimated by the mass function. Because the precise probability distribution of C may not be known exactly, in evidential theory, bounds of probability distribution are defined. The lower and upper probability of a subset B of C are denoted as B's belief measure $\text{Bel}_m(B)$ and plausibility measure $\text{Pls}_m(B)$, respectively. They can be determined from the mass function as follows:

$$\text{Bel}_m(B) = \sum_{A \subset B} m(A) \qquad (2)$$

$$\text{Pls}_m(B) = \sum_{A \cap B \neq \Phi} m(A) \qquad (3)$$

Generally, $\text{Bel}_m(B) \neq \text{Pls}_m(B)$ and, therefore, somewhere in the belief interval $[\text{Bel}_m(B), \text{Pls}_m(B)]$ lies the true probability of B. In evidential theory, $\text{Bel}_m(B)$ indicates the amount of belief committed to B based on the given piece of evidence, while $\text{Pls}_m(B)$ represents the maximum extent to which the current evidence allows one to believe A.

If m_1 and m_2 are two mass functions of C induced by

TABLE 1. BASIC PROBABILITY ASSIGNMENT VALUES FOR A SET OF EVIDENCE FROM TWO SPECTRAL BANDS FOR THE CLASSIFICATION OF FOUR CLASSES

	Urban (U)	Agriculture (A)	Forest (F)	Water (W)
Band 1	0.2	0.3	0.3	0.1
Band 2	0.1	0.3	0.4	0.0

two mapping, $\Gamma_1: E_1 \rightarrow C$ and $\Gamma_2: E_2 \rightarrow C$, where E_1 and E_2 are independent sources of evidence, then the combined mass function, denoted by $m_1 \oplus m_2$, can be calculated using Dempster's rule of combination: i.e.,

$$m_1 \oplus m_2 (D) = \frac{\sum_{A_i \cap B_j = D} m_1(A_i) m_2(B_j)}{1 - \sum_{A_k \cap B_j = \Phi} m_1(A_k) m_2(B_j)} \quad (4)$$

where the combination operator "$\oplus$" is called "orthogonal sum," $D \subset C$ and $D \neq \Phi$, and $m_1 \oplus m_2 (\Phi) = 0$. Using the orthogonal summation, one can update the beliefs and plausibilities in space C with additional sources of evidence. If a piece of evidence from a third source is given, we can treat $m_1 \oplus m_2$ as m_1 or m_2 and apply m_3 in the same manner as we combine m_1 and m_2. Because operator "$\oplus$" is commutative and associative, the order of applying the orthogonal summation does not affect the final results.

A number of applications of the evidential theory can be found in expert system development (e.g., Gordon and Shortliffe, 1985; Shafer and Logan, 1987; Kruse et al., 1991) and the applications of knowledge-based systems to image analysis (Goldberg et al., 1985; Srinivasan and Richards, 1990; Kontoes et al., 1993).

To illustrate the orthogonal sum, consider data from two spectral bands as two independent sources for classification. The "BPA" values from the mass functions are listed in Table 1. None of the row-wise sums in Table 1 equals 1. The residual, $[1 - m(U) - m(A) - m(F) - m(W)]$, treated as the degree of ignorance, is denoted by I. To calculate $m_1 \oplus m_2(D)$, where $D \subset C = \{U, A, F, W\}$, we illustrate the procedure using Table 2. Table 2 is divided into two parts, the top part is used for calculating the mass product and the second part is devoted to the orthogonal sum and the beliefs and plausibilities.

A requirement for use of evidential theory is that evidence sources are independent of each other (Shafer, 1976). This independence is, however, not solely a statistical one. Two highly correlated sources of data may seem redundant in a statistical sense but can improve our confidence of the classification results obtained from evidential reasoning. For the sake of simplicity and the lack of a way of verifying evidential independence, researchers often use all sources available (Lee et al., 1987; An et al., 1992; Peddle, 1993). Alternatives to reducing statistical dependencies are (1) to treat each individual source, E_i, as a component in a subspace of E rather than an independent source, and (2) to decorrelate multisource data through principal component analysis or factor analysis (Durrand and Kerr, 1989).

To apply the evidential theory to a classification problem, the following steps are needed:

(1) determine the probability distribution $p_{ij}(x_i)$ of C_j for each evidential source E_i (note that this is different from algorithms based on Bayes theory in which construction of a mapping between the entire feature space E to an individual class C_j is usually required). Each band of an image or a digital map can be considered as an individual evidential source. For example, the histogram of the "ith" image can be used to approximate a probability distribution denoted by $p_{ij}(DN)$, where $DN \in \{0, 1, ..., 255\}$ for an 8-bit image;
(2) determine the mass function of each class C_j, $j = 1, ..., k$, for each evidential source E_i (Yen, 1989): i.e.,

$$m_i(C_j) = \sum_{f(DN) = C_j} p_{ij}(DN)$$

where $f(DN) = C_j$ defines a mapping between value DN in evidential source E_i and class C_j. If C_j is a singleon class, $m_i(C_j) = p_{ij}(DN)$;
(3) combine the mass functions from two evidential sources using Equation 4. This formula can be used iteratively, one source at a time, until "BPA" values from all sources are combined;
(4) determine the belief interval for each class C_j. Assuming that the combined mass function, m, has been obtained from Step 3, the belief interval can be calculated using Equations 2 and 3; and
(5) base the classification of a set of evidence or observations and measurements, $\mathbf{X} = (x_1, x_2, ..., x_n)$, on either the total belief or the total plausibility (Lee et al., 1987).

Different from classification algorithms based on Bayes theory (see Gong and Dunlop, 1991), algorithms based on evidential theory are restricted neither by the dimension of the spatial data nor by the number of data sources. Therefore, a

TABLE 2. COMPONENTS FOR CALCULATING THE DEMPSTER'S ORTHOGONAL SUM

Band 1 Band 2		U 0.2	A 0.3	F 0.3	W 0.1	I 0.1
U	0.1	U 0.02	Φ 0.03	Φ 0.03	Φ 0.01	U 0.01
A	0.3	Φ 0.03	A 0.09	Φ 0.09	Φ 0.03	A 0.03
F	0.4	Φ 0.08	Φ 0.12	F 0.12	Φ 0.04	F 0.04
I	0.2	U 0.04	A 0.06	F 0.06	W 0.02	I 0.02
		$\sum_{G \cap H = U} m_j(G) \cdot m_2(H) = 0.07$	$\sum_{G \cap H = A} m_j(G) \cdot m_2(H) = 0.18$	$\sum_{G \cap H = F} m_j(G) \cdot m_2(H) = 0.22$	$\sum_{G \cap H = W} m_j(G) \cdot m_2(H) = 0.02$	$\sum_{G \cap H = I} m_j(G) \cdot m_2(H) = 0.02$
		$1 - \sum_{G \cap H = \Phi} m_j(G) \cdot m_2(H) = 0.51$				
$m_1 \oplus m_2$		0.07/0.51 = 0.14	0.18/0.51 = 0.35	0.22/0.51 = 0.43	0.02/0.51 = 0.04	0.02/0.51 = 0.04
$Bel_{m_1 \oplus m_2}$		0.14	0.35	0.43	0.04	0.04
$Pls_{m_1 \oplus m_2}$		0.18	0.39	0.47	0.08	0.04

non-parametric model can be employed with no dimensionality limitation. Because the lack of certain data sources only reduces the number of times that Equation 4 is employed, to some extent the evidential reasoning algorithm is tolerant to incomplete data coverage. Moreover, the evidential reasoning algorithm does not differentiate the *a priori* and *a posteriori* probabilities. It does not require any *a priori* probabilities to be known explicitly. Therefore, the evidential reasoning algorithm has fewer limitations as compared with the Bayesian algorithms.

Classification Using a Feed-Forward Neural Network

A network of elemental processors arranged and connected in a feed-forward manner reminiscent of biological neural nets can be used to classify a set of observations, $\mathbf{X} = [x_1, x_2, ..., x_n]^T$ from n different sources, and label it into a class $C_i \in C = \{C_1, C_2, ..., C_n\}$. Rumelhart *et al.* (1986) developed a generalized delta rule (GDR) for supervised training of a neural network based on error back propagation.

The architecture of a layered net with feed forward capability is shown in Figure 1. The basic elements are nodes "$\bigcirc$" and links "$\rightarrow$". Nodes are arranged in layers and each of them is a processing element. Each input node accepts a single value which corresponds to an element in $\mathbf{X}$. Each node generates an output value. Depending on the layer in which a node is located, its output may be used as the input for all nodes in the next layer. The links between nodes in successive layers are weight coefficients. The number of hidden layers can be greater than 1. In the output layer, each node corresponding to a single class in C generates the membership value v_k of that class. For example, ω_{ji} is the link between two nodes from layer i to its successive layer j. Each node, except those in the input layer, is an arithmetic unit. It takes the inputs from all the nodes of its previous layer and uses the linear combination of those input values as its net input. For a node in layer j, the net input is

$$u_j = \sum \omega_{ji} \cdot x_i .$$

The output of the node in layer j is

$$O_j = f(u_j)$$

where f is an activation function that often takes the form of a sigmoidal function,

$$O_j = \frac{1}{1 + e^{-(u_j + \theta_j)}} \qquad (5)$$

where θ_j serves as a threshold or bias. This function allows for each node to react to an input differently. Some nodes may be easily activated or fired to generate a high output value when θ_j is large. On the contrary, when θ_j is small, a node will have a slower response to the input u_j. This is considered occurring in the human neural system where neurons are activated by different levels of stimuli.

Such a feed-forward network requires a single set of weights and biases that will satisfy all the input-output pairs presented to it. The input is a set of observations and the outputs are the desirable class membership values $\mathbf{V}_p = \{v_{p1}, v_{p2}, ..., v_{pk}\}$. The process of obtaining the weights and biases is network learning, which is essentially the same as supervised training. During network training, elements in a set of observations $\mathbf{X}_p = \{x_{p1}, x_{p2}, ..., x_{pm}\}$ correspond to the nodes in the input layer. For the given input $\mathbf{X}_p$, we require that the network adjust the set of weights in all the connecting links and also all the thresholds in the nodes such that the desired outputs can be obtained. Once this adjustment has been accomplished by the network, another pair of $\mathbf{X}_p$ and $\mathbf{V}_p$ is presented and the network is asked to learn that association also. In general, the output from the net $\mathbf{O}_p = \{o_{pq}\}$ will
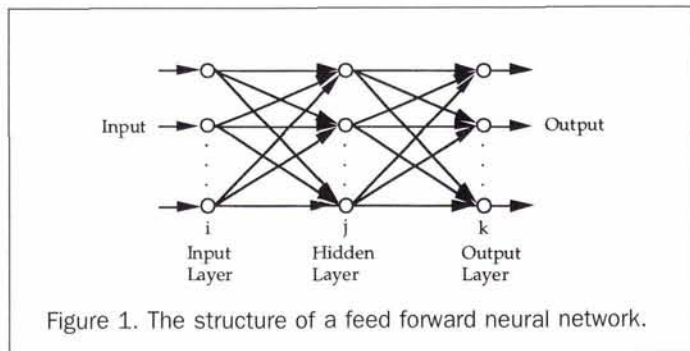

Figure 1. The structure of a feed forward neural network.

not be the same as the desirable values $\mathbf{V}_p$. For each $\mathbf{X}_p$, the squared error is

$$\varepsilon_p = \sum_{q=1}^{k} (v_{pq} - o_{pq})^2 \qquad (6)$$

where k is the number of classes and the average system error is

$$\varepsilon = \frac{1}{nt} \sum_{p=1}^{nt} \sum_{q=1}^{k} (v_{pq} - o_{pq})^2$$

where nt is the number of training pairs.

The adjustment of weights and thresholds is accomplished by repetitively feeding the network with the $\mathbf{X}$ and $\mathbf{V}$ pairs in sequence and constantly modifying the weights and thresholds using the generalized delta rule (GDR). With GDR, the correct set of weights is obtained by varying the weights in a manner calculated to reduce the error ε_p as rapidly as possible. In general, different results will be obtained between the use of ε_p and ε during the training based on error back propagation (Pao, 1989).

The convergence of ε_p with improved values of weights and thresholds is achieved by taking incremental changes that are proportional to the partial derivatives from Equation 6. For weight adjustment, this is done by modifying weights with an increment proportional to $-\partial E_p / \partial \omega_{ji}$, i.e., with an adjustment of $\mu(-\partial E_p / \partial \omega_{ji})$. Starting at the output layer, GDR propagates the "error" backward to earlier layers. Thresholds θ_j are learned in the same manner as are the weights. Parameter μ is a small positive number experimentally determined and usually fixed during each training process. Our experiences suggest that large differences in range from one data source to another make it harder for us to select μ. When the input data are converted to the range of [0, 1], it is easier to find an appropriate μ so as to make the network learn faster. Data range conversion can be achieved by finding the maximum and minimum in each channel and applying the following linear transformation to the original data:

new data value =
(original data − minimum)/(maximum − minimum).

This is similar to data normalization as suggested in Azimi-Sadjadi *et al.* (1993) and Freeman and Skapura (1991). Details on the learning algorithm can be found in various texts (e.g., Pao, 1989; Freeman and Skapura, 1991). Although a three-layer network can form arbitrarily complex decision regions, sometimes difficult learning tasks can be simplified by increasing the number of internal layers (Pao, 1989). On the other hand, were too many layers in a network or too many nodes in a layer used, the network would require much more computation and might lose the ability to generalize. Because feed-forward network nodes in the same layer

are independent of each other, they can be implemented in parallel processing.

Algorithms similar to the one explained above have been applied to land-cover classification of remote sensing data only (Hepner et al., 1990; Civco, 1993; Dreyer, 1993). Other neural network algorithms have also been developed and applied to remote sensing image classification (e.g., Benediktsson et al., 1993; Salu and Tilton, 1993). A detailed review is found in Sui (1994).

Discussion of the Algorithms
In classification, expert knowledge on the spatial location of classes is used to train the algorithms. To do so, we transform this type of knowledge into a computer system. The processes of collecting and encoding expert knowledge are referred to as knowledge acquisition and knowledge representation, respectively. While various complex computer structures for knowledge representation may be used, relatively simple procedures are often employed such as the use of parametric statistical models (Swain and Davis, 1978; Jensen, 1986; Richards, 1986) or non-parametric look-up tables (Duda and Hart, 1973).

Training in the Evidential Reasoning Algorithm
Moon (1990) intuitively assigned probabilities based on expert knowledge and suggested that a more systematic and quantitative approach be established. Shi (1994) used a combination of both parametric and non-parametric modeling in the construction of mass functions. He proposed the following methods:

- *Occurrence-Frequency Table.* Occurrence frequency densities can be estimated for data that are obtained from any measurement scales. For data at nominal or ordinal measurement scales, this is the only method that can be used. The frequency table is built with its row entries being the values of what a data source may have and each of its columns corresponding to a class; and
- *Normal Distribution Model.* Similar to MLC, a normal distribution model can be built for each class for those data sources that are acquired at the interval and ratio scales.

If the occurrence-frequency table method is applied to data at interval and/or ratio measurement scales, certain interpolation or extrapolation methods can be used to adjust the occurrence table to fill some of the gaps in the feature space caused by under-sampling in training samples. When training samples are too few or do not exist, one may use fuzzy set theory to establish mass functions. With fuzzy set theory (Zadeh, 1965), expert knowledge can be encoded using fuzzy membership functions (Liu and Burrough, 1986; Mulder and Corns, 1993; Zhu and Band, 1994). To do so, one needs to determine the fuzzy membership function on each source E_i, $i = 1, 2, ..., n$ for each C_j, $j = 1, 2, ..., k$. Thus, a total of n by k membership functions need to be found. Fuzzy membership functions can then be normalized or transformed using other methods so as to meet the requirements of mass functions. It should be noted that one of the advantages of evidential theory is that it allows training to be conducted in hierarchical classification problems. This facet, however, will not be explored here.

Training in the Neural Network Algorithm
In contrast to the evidential theory based algorithm, the GDR training process in the neural network algorithm encodes knowledge through its weights and thresholds associated with each node on the net. Explicit modeling of data sources is not required in the neural network method. In addition, there is no need to treat the data sources independently (Benediktsson et al., 1993).

The training process is computationally intensive, however. It requires many training samples and many iterations and it is usually terminated when the system error calculated from Equation 6 is smaller than a preset value. We can monitor the progress of training periodically by applying feed-forward calculation through the network to classify the training samples as well as some independent test samples. System errors calculated for the training samples and testing samples can be plotted against the number of iterations for the purpose of checking the performance of the network.

Uncertainty Measures
An advantage of the use of evidential theory as compared to the use of probability is its ability to express ignorance. The commitment of belief to a subset B does not force the remaining belief to be committed to its complement, i.e., $\text{Bel}_m(B) + \text{Bel}_m(\overline{B}) \leq 1$. The amount of belief committed to neither B nor B's complement is the degree of ignorance.

For neural network techniques, uncertainty about a classification result can only be obtained when the first type of classification, as discussed in the section on Classification, is to be made. For example, for a given vector $\mathbf{X}$, with evidential reasoning we can measure the uncertainties in membership degrees of individual classes in C. With neural network methods, only when $\mathbf{X}$ is classified into a particular class can the uncertainty involved in that class be determined.

Experiments
To illustrate the classification algorithms based on evidential theory and neural networks for integrated analysis of multisource data, we selected a data set for geological mapping in Melville Peninsula, Northwest Territories, Canada. The algorithms have been implemented using C programming language on SUN Workstations in the Remote Sensing Laboratory of the Department of Geomatics Engineering, The University of Calgary.

Study Site and Multisource Spatial Data
The study site is a 12- by 12-km area centered at approximately 68°32' N and 82°43' W. This area, located west of Hall Lake on Melville Peninsula in the northeast Arctic of Canada, has relatively large areas of unweathered outcrop and relatively little vegetation. It has been studied for predicting geological units from Landsat Thematic Mapper (TM) and geophysical data (Chung et al., 1993).

The multisource spatial data set used in this study was processed and provided by the Geological Survey of Canada. It includes Landsat TM, gamma-ray spectrometer, magnetometer, and gravity anomaly data. While the TM data contain information primarily about the surface, the other three types of data reveal properties of rock materials for up to 1 m, 1 km, and 10 km below the surface, respectively. The radiometric data from the gamma-ray spectrometer were originally acquired along flight lines at a spacing of 5 km, with a sampling frequency of approximately 130 m along each flight line. Four types of gamma-ray radiometric data were obtained, including total exposure, potassium, uranium, and thorium. The data were subsequently resampled to 500- by 500-m grids. The aeromagnetic data were originally acquired along flight lines at a spacing of 1 km with a sampling frequency of approximately 80 m. The data were subsequently resampled to 200- by 200-m grids. The gravity data were originally acquired from ground stations of approximately 10 km^2. They were resampled to 2- by 2-km grids. In addition, a surficial geology map was digitized for geological mapping purposes.

All the different sources of data were geometrically transformed to the UTM coordinate system and resampled to the TM image pixel size of 30 m by 30 m. It should be noted

that resampling of data from a coarser resolution to a finer one does not increase the information level. Although it may increase the amount of computation, resampling simplifies the application of various classification algorithms. The data set containing 13 channels can thus be used as input to the two algorithms. Figure 2 shows the Landsat TM Band 7 image of 400 by 400 pixels (Figure 2a), the potassium-radiometric image (Figure 2b), the aeromagnetic data (Figure 2c), and the gravity anomaly data (Figure 2d). A map of outcrop areas of four different types of geological classes was derived by intersecting the digitized surficial geology map and a vegetation index map from the Landsat TM data (Chung *et al.*, 1993). The four geological classes of outcrops are (1) Tonalite gneiss, (2) Prince Albert group, (3) Hall Lake plutonic complex, and (4) Ordovician carbonates (see Table 3).
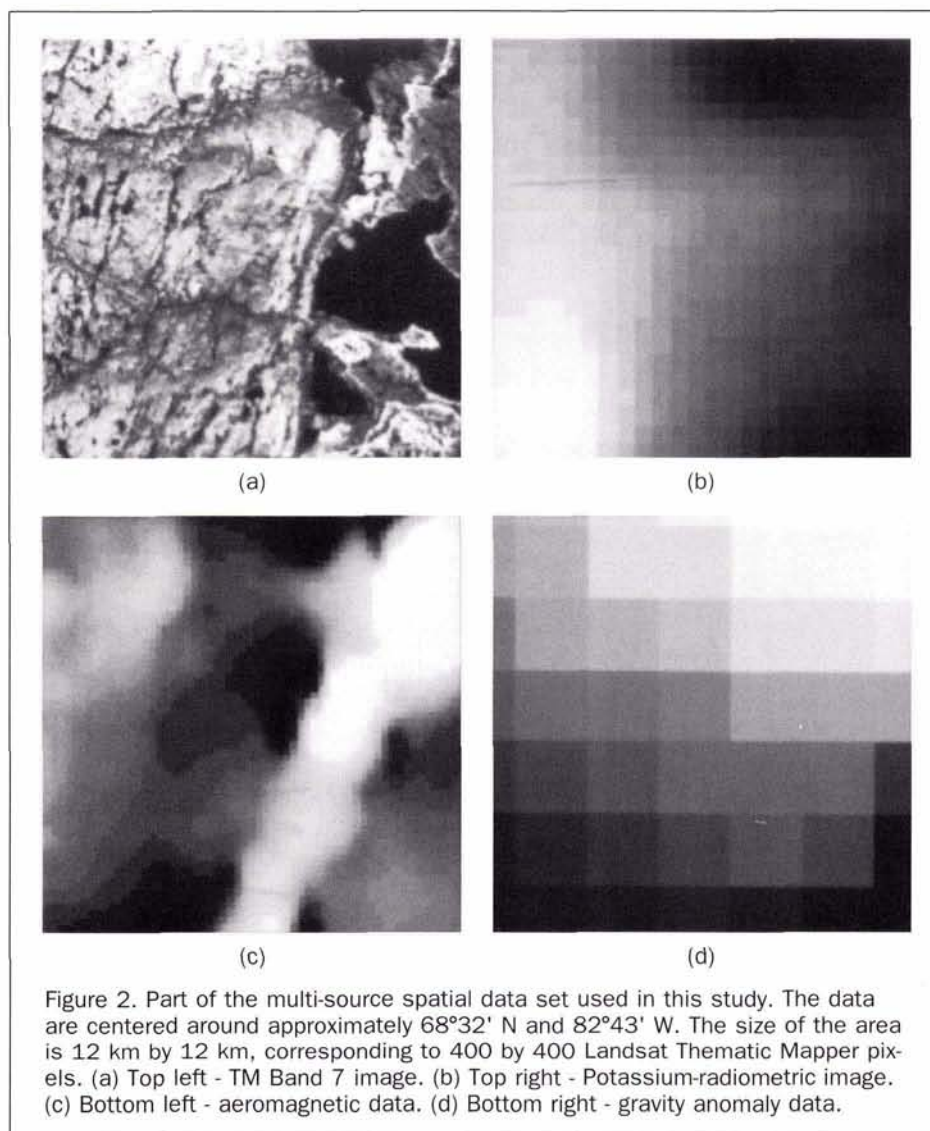
Training and Testing the Algorithms

Samples for training and testing were selected from completely non-vegetated areas as indicated by low vegetation indices calculated from the red and near-infrared bands of the Landsat TM data (Chung *et al.*, 1993). The total number of sample pixels for training and testing is listed in Table 3 for each class. From the total samples of each class, training samples were randomly selected at a specific percentage and

TABLE 3. TOTAL NUMBER OF SAMPLES FOR EACH GEOLOGICAL CLASS AND PERCENTAGES USED FOR TRAINING

Geological Class	Total Number of Samples	Samples Used for Training the Evidential Algorithm (%)	Samples for Training the Neural Network (%)
1. Tonalite gneiss	468	10.3 (22.4)*	13.0
2. Prince Albert group	583	10.1 (23.2)*	12.7
3. Hall Lake plutonic complex	3678	10.8 (26.2)*	9.6
4. Ordovician carbonates	101	7.0 (39.6)*	24.8
Total Samples/ Overall %	4830	10.6 (25.7)*	10.6

*Percentage of the second set of training samples in bracket

all the remaining samples were used as test samples. This ensured that training and test samples were not overlapping. In doing so, we expected that this would cause less bias and more representative estimation of classification accuracies. After each classification, we compared the classification re-

(a)

(b)

(c)

(d)

Figure 2. Part of the multi-source spatial data set used in this study. The data are centered around approximately 68°32' N and 82°43' W. The size of the area is 12 km by 12 km, corresponding to 400 by 400 Landsat Thematic Mapper pixels. (a) Top left - TM Band 7 image. (b) Top right - Potassium-radiometric image. (c) Bottom left - aeromagnetic data. (d) Bottom right - gravity anomaly data.

sults with the test samples to generate a confusion matrix based on which classification accuracies were assessed (Story and Congalton, 1986).

For the evidential reasoning method, two sets of training samples were selected. The percentages of training samples for each class are listed in Table 3. The first set of training samples contains approximately 10.6 percent of the total sample, and the percentage for each class is listed in Table 3 without brackets. The percentages in the brackets represent the second set of training samples, which account for approximately 25.7 percent of the total sample. The idea was to test the sensitivity of the evidential reasoning algorithm against the size of training samples. Generally, we prefer using training samples that are small in size, provided that the training samples are representative.

For the neural network algorithm, approximately 10.6 percent of the total sample (Table 3) was randomly selected for training. After the training of the neural network, the entire data set was used as input to the neural network and the output was compared with the remaining 89.4 percent of total samples to generate the confusion matrix.

Test of the Evidential Reasoning Algorithm

Parametric models and lookup tables can be used to estimate "BPA" during supervised training. While the data set used in this study does not have any restriction on the use of either method, a lookup table for each channel of data was considered essential because of its simplicity. It was subsequently realized that, because the surface radiance variability in the TM bands was very large with respect to each geological class, the use of lookup tables with TM bands caused much confusion among classes. We tried to apply a Gaussian model to the seven TM bands. Therefore, a combination of Gaussian parametric models associated with the TM bands and lookup tables with the rest of the channels was tested also. Based on the two different ways of "BPA" construction and the use of two sets of training samples, we obtained four sets of classification results. The confusion matrices for these four classifications were generated (Tables 4 through 7). Tables 4 and 5 correspond to the classification results obtained by constructing the "BPA" from the 10.6 percent training sample set with

- the results obtained with a combination of the Gaussian model and lookup tables shown in Table 4, and
- the results obtained from the use of lookup tables only shown in Table 5.

Tables 6 and 7 correspond to the classification results obtained by constructing the "BPA" from the 25.7 percent training set with

- the results obtained from a combination of the Gaussian model and lookup tables presented in Table 6, and
- the results obtained from lookup tables only shown in Table 7.

All confusion tables are arranged with the classified classes as column entries and the test classes as row entries. The off-diagonal elements along a row indicate the number of test samples in that class omitted by the classification algorithm. We derived a producer's accuracy for each row and used it as an accuracy measure for each individual class (Story and Congalton, 1986). We also calculated the Kappa coefficient and its variance from each confusion table. The Kappa coefficient is an overall classification accuracy measure that excludes chance agreement (Cohen, 1960). The variances and the Kappa coefficients calculated from two confusion matrices were used to derive a Z-value for testing if two classification accuracies are significantly different (Fleiss et al., 1969).

TABLE 4. CLASSIFICATION RESULTS OBTAINED USING THE EVIDENTIAL REASONING ALGORITHM WITH THE COMBINATION OF PARAMETRIC MODELS AND LOOKUP TABLES OBTAINED FROM THE 10% TRAINING SAMPLES

Class	Classification				Producer's Accuracy
	1	2	3	4	
1	303	23	94		72.1%
2	9	427	88		81.5%
3		43	3230	7	98.5%
4		9	7	78	83.0%
Overall					93.5%
Average					83.8%
Kappa Coefficient					0.828
Variance of Kappa					0.000183

TABLE 5. CLASSIFICATION RESULTS OBTAINED USING THE EVIDENTIAL REASONING ALGORITHM WITH THE LOOKUP TABLES OBTAINED FROM THE 10% TRAINING SAMPLES

Class	Classification				Producer's Accuracy
	1	2	3	4	
1	104	2	36	278	24.8%
2	43	76	59	346	14.5%
3	20	10	3131	118	95.5%
4	52	14	3	25	26.6%
Overall					77.3%
Average					40.4%
Kappa Coefficient					0.459
Variance of Kappa					0.000712

TABLE 6. CLASSIFICATION RESULTS OBTAINED USING THE EVIDENTIAL REASONING ALGORITHM WITH THE COMBINATION OF PARAMETRIC MODELS AND LOOKUP TABLES FROM THE 25.7% TRAINING SAMPLES

Class	Classification				Producer's Accuracy
	1	2	3	4	
1	294	6	63		81.0%
2	15	353	80		78.8%
3	3	20	2691	1	99.1%
4			1	60	98.4%
Overall					94.7%
Average					89.3%
Kappa Coefficient					0.861
Variance of Kappa					0.000162

TABLE 7. CLASSIFICATION RESULTS OBTAINED USING THE EVIDENTIAL REASONING ALGORITHM WITH THE LOOKUP TABLES FROM THE 25.7% TRAINING SAMPLES

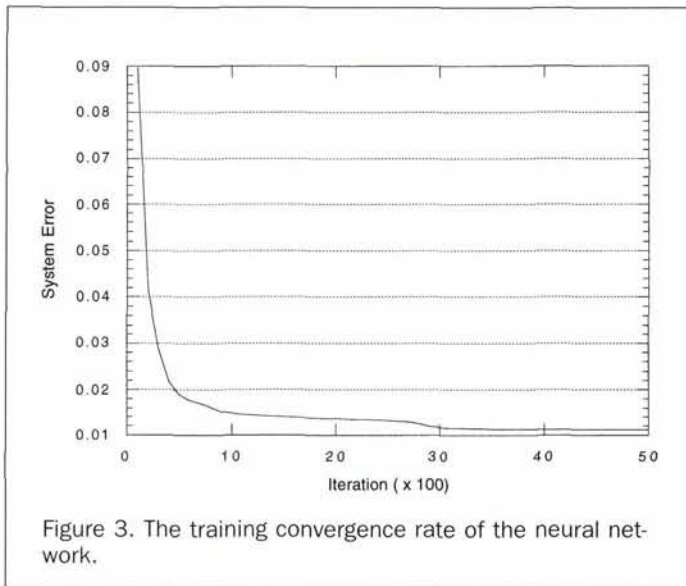| Class	Classification				Producer's Accuracy
	1	2	3	4	
1	71	18	33	241	19.6%
2	40	48	76	284	10.7%
3	14	8	2633	59	97.0%
4	18	11	2	30	49.2%
Overall					77.6%
Average					44.1%
Kappa Coefficient					0.454
Variance of Kappa					0.000927

Figure 3. The training convergence rate of the neural network.

A Z-value of greater than 1.96 indicates that the difference between two classification accuracies is statistically significant at the confidence level of 0.95 probability.

Test of the Neural Network Algorithm
The neural network algorithm based on error back propagation using the GDR was adapted from Pao (1989). Three layers were used to construct the network for the classification. The number of nodes in the input layer was 13, corresponding to the 13 channels of the multisource data set. There were four nodes in the output layer each corresponding to a geological class. Input data were linearly converted to the range of [0, 1].

In the training data, each input vector originally has a class assigned to it as the desirable output. This class has to be converted to a vector through binary one-of-n-coding for use with the neural network algorithm. For example, if the desirable class is geological class 2. The output vector required by the network should be [0, 1, 0, 0], assuming that the second output node corresponds to the second class. As can be seen from Equation 5, it takes a longer time for a network to generate binary outputs such as [0, 1, 0, 0]. To speed up the training process, we used [0.003, 0.99, 0.003, 0.003] instead. At the classification stage, the decision on the classification of the input vector is made based on the highest value in the output vector.

The number of nodes in the hidden layer has to be determined empirically. We started with 100 nodes to test the training convergence and subsequently reduced to 26 nodes, which is two times the input nodes, and the network converged well. With the network structure of 13 input nodes, 26 hidden nodes, and four output nodes, less than 500 iterations were required to reach a low level of system error, 0.02, for the training data, and the trend had stabilized since then (Figure 3). We terminated the training process after 5000 iterations. The weights and thresholds obtained were applied to classify the entire data set. The test results are summarized in Table 8.

Discussions
From Table 3, we can immediately find out the imbalance of population among the four classes. Class 3 accounts for over 75 percent of the total population. While classes 1 and 2 each have about 10 percent of the overall population, the population of class 4 is very small. Were the overall classifi-

cation of our primary concern, as in some classification tasks such as land-cover or land-use classifications, class 4 would not attract much of our attention because of its small population. In geological mapping, however, a class with a small population may contain more valuable mineral deposits than would other classes. The importance of a geological class *should therefore not be decided based on its size*. This implies that the average of individual class accuracies is perhaps more important than an overall classification accuracy for the purpose of this study.

By comparing Tables 4 and 6 with Tables 5 and 7, we can see that for the evidential reasoning the use of a combination of parametric models and lookup tables in constructing the "BPA" has produced considerably more accurate results than has the use of lookup tables only. This is particularly true for classes 1, 2, and 4. Class 3, with the largest population among the four classes, has been classified very well, because the lowest accuracy for this class even reaches 95.5 percent in Table 5. It seems that the first training strategy is superior to the second one.

As we compare Figure 4a and Figure 4b, however, Figure 4b looks surprisingly more heterogeneous than does Figure 4a. The noisy appearance of Figure 4b results from the strong influence of TM images because all the other types of data have a much coarser spatial resolution. The non-normality and large variability for each class, as characterized by the TM image, probably contribute very little to the actual discrimination of one class from another. The "BPA" functions obtained by applying the lookup table method to the TM bands have magnitudes comparable to those obtained from the other types of data. This led to the variability from the TM data being carried through the iterative orthogonal summation to the end product as manifested in Figure 4b. On the other hand, the large variability caused the large standard deviations in the geological classes which led to low magnitudes in the resultant "BPA" functions modeled by the Gaussian distribution. Consequently, the low "BPA" values from Gaussian models have been buried by "BPA" values from other data types through the iterative orthogonal summation. This resulted in Figure 4a being less noisy. The higher classification accuracies in Figure 4a should be interpreted as an indication that the TM images contributed little to this particular geological classification. From the above discussion, it is clear that the lookup table method for constructing "BPA" functions is sensitive to data variation and vulnerable to noise. If certain data types are noisy, the use of a parametric model may suppress the noise level. Therefore, a combination of the parametric modeling approach with the lookup table method can be an effective tool in constructing "BPA" functions.

Because the dominance of class 3, the increase of train-

TABLE 8. CLASSIFICATION RESULTS OBTAINED USING THE NEURAL NETWORK ALGORITHM

	Classification				Producer's Accuracy
Class	1	2	3	4	
1	401	6			98.5%
2	14	405	73	17	79.6%
3		47	3269	9	98.3%
4	2	6		68	91.9%
Overall					96.0%
Average					92.1%
Kappa Coefficient*					0.894
Variance of Kappa					0.000092

*The Kappa Coefficient of 0.894 represents a statistically significant improvement at 0.95 probability confidence level over the 0.861 calculated from Table 6.
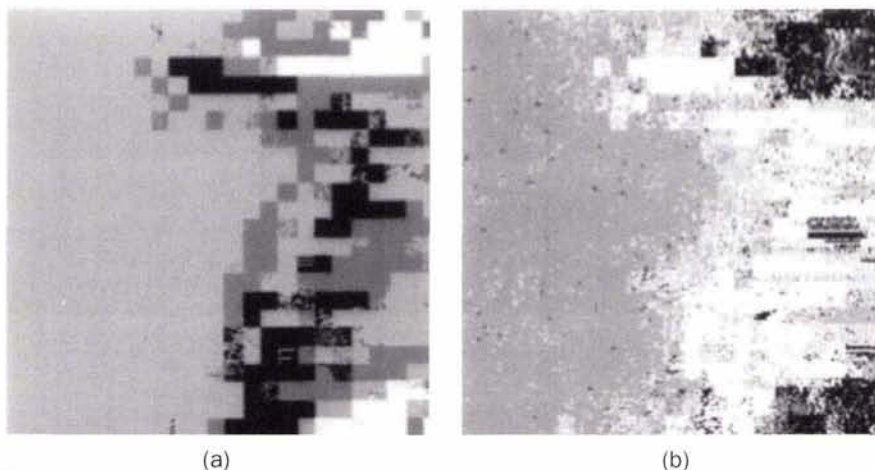
Figure 4. Evidential reasoning classification results from the entire data set. (a) Left - the basic probability assignments are obtained using a combination of parametric modeling and lookup tables, i.e., Gaussian distribution models have been used for the seven TM channels and lookup tables have been applied to the remaining channels of the data set. (b) Right - the basic probability assignments are obtained solely from lookup tables for all the channels. Black - Tonalite gneiss; Dark Grey - Prince Albert group; Light Grey - Hall Lake plutonic complex; and White - Ordovician carbonates.

ing sample size from approximately 10 percent to approximately 25 percent of the total samples has not led to a significant improvement in the overall accuracies (1.2 percent difference between Tables 4 and 6, and 0.3 percent difference between Tables 5 and 7). Larger differences, however, are observable among the average accuracies. For example, the increase in training sample size has resulted in an average accuracy improvement of 5.5 percent between Tables 4 and 6. When comparing each individual class in Tables 4 and 6, we can see that, except for a 2.7 percent drop in class 2, there are accuracy improvements for three of the classes, especially for classes 1 and 4, as the training sample sizes increase. However, when Tables 5 and 7 are compared class by class, classes 1 and 2 have 5.2 percent and 3.8 percent drops in accuracy, respectively. We suspect again that it is the noisiness of the TM data that affects the improvement of accuracies with these classes. We have attempted to increase further the training sample sizes to 75 percent of the total sample but no significant accuracy improvements were achieved.

Comparing the classification results (Tables 4 and 6) obtained from the evidential reasoning with those from the neural network algorithm (Table 8), we can see that the best overall accuracy of 96.0 percent and the best average accuracy of 92.1 percent have been achieved with the neural network algorithm. When a similar number of training samples was used (Table 4 versus Table 8), accuracy improvements by the neural network are 2.5 percent and 8.3 percent, respectively, for the overall and average accuracies. When the chance agreement is removed by comparing the Kappa coefficients in Tables 4 and 8, the accuracy improvement is 0.066. The accuracy improvement by the neural network algorithm in comparison with the evidential reasoning algorithm is statistically significant at the 0.95 probability confidence level. When we compare Tables 6 and 8, there is only a 1.3 percent overall accuracy improvement achieved with the neural network algorithm over the evidential reasoning algorithm. The average accuracy improvement is 2.8 percent. The accuracy improvement measured by Kappa coefficient of 0.033, however, is still significant at the 0.95 probability confidence level.

As the producer's accuracies of individual classes are examined among Tables 4, 6, and 8, with the exception of class 1 (98.5 percent in Table 8), the best accuracies for classes 2 (81.5 percent in Table 4), 3 (98.5 percent in Table 6), and 4 (98.4 percent in Table 6) have all been achieved with the evidential reasoning. When Figure 5 is compared with Figure 4, we can see that the classification results obtained from the neural network are very unique. On the one hand, the classification results in Figure 5 contain much less noise as compared with Figure 4b. This indicates that the neural network algorithm is less sensitive to noise than is the lookup table method used with the evidential reasoning. On the other hand, the class boundaries in Figure 5 look more
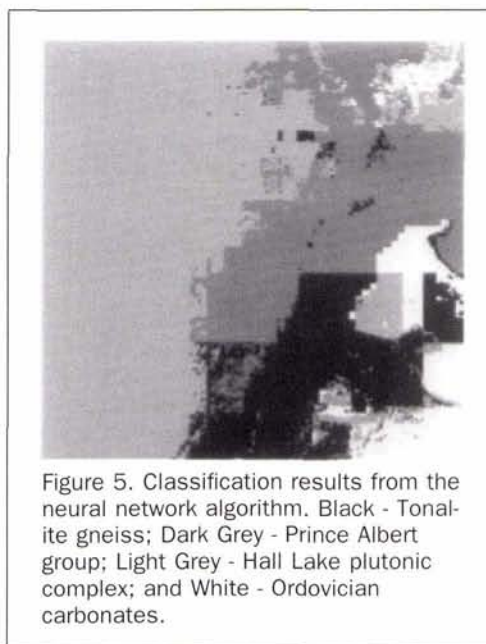

Figure 5. Classification results from the neural network algorithm. Black - Tonalite gneiss; Dark Grey - Prince Albert group; Light Grey - Hall Lake plutonic complex; and White - Ordovician carbonates.

natural and similar to the geological map made manually than do those in Figure 4a. It seems that the neural network method is more adaptive and more powerful in generalization than the evidential reasoning method.

From a computation point of view, the neural network algorithm consumed more time during the training stage than the time required by the entire classification with the evidential reasoning method. However, the evidential reasoning method requires more intervention. For instance, it requires one to find out how the variability of each data source may affect the specific classification task. Much time needs to be spent on learning how to construct the "BPA" functions. In fact, this is the most critical step involved in the use of the evidential reasoning algorithm. In this experiment, we spent more time in testing the evidential reasoning algorithm than we did in applying the neural network algorithm.

Summary and Conclusions
A long term goal of our studies is to develop algorithms for integrated analysis of spatial data from multiple sources. This study was to compare evidential reasoning with error back-propagation feed-forward neural network methods for classification of multisource spatial data. The evidential reasoning and the neural network methods were tested using a data set composed of Landsat TM data, gamma ray radiometric data, aeromagnetic data, and gravity anomaly data acquired from an area in Melville Peninsula, Northwest Territories, Canada. The data set was used to classify four geological classes.

With the evidential reasoning algorithm, we used the lookup table method and Gaussian model to construct "BPAs." "BPAs" based on lookup tables from all data sources generated poor classification results. Much better results were achieved when Gaussian modeling was used to construct "BPAs" for the Landsat TM data while lookup tables were used for the remaining data sources. The improvements for average accuracies exceeded 40 percent with the application of Gaussian modeling to the Landsat data. We believe that the Landsat TM data contribute marginally to the classification of the four geological classes. The evidential reasoning method based on the lookup tables is sensitive to data noise and large data variability. The experiment suggests that parametric modeling can be used to complement the table lookup method. Other methods such as texture analysis and filtering that can be used to preprocess the noisy data may also prove helpful.

The results indicate that selection of an appropriate training strategy is most critical for successful use of evidential reasoning methods. Because selecting training strategies is essentially a task of knowledge acquisition and encoding, we believe that knowledge acquisition and encoding play an equivalently important role in expert system-based decision making.

With a small proportion of training samples (10.6 percent, and the smallest contained only seven samples for one class!), the evidential reasoning method generated reasonable results with an overall accuracy of 93.5 percent and an average accuracy of 83.8 percent. An overall accuracy of 94.7 percent and average accuracy of 89.3 percent were achieved when the training samples accounted for approximately a quarter of the total sample population.

The best overall accuracy of 96.0 percent and average accuracy of 92.1 percent were achieved using a single-hidden-layer neural network. Based on Kappa coefficients calculated, the overall accuracy of neural network classification is significantly higher than those accuracies obtained with the evidential reasoning algorithm. The network structure contained 13, 26, and 4 nodes in the input, hidden, and output layers, respectively. The neural network, however, only generated one of the four highest accuracies for individual classes when compared with the evidential reasoning algorithm. Because accurate classifications for certain individual classes are highly desirable in geological mapping, evidential reasoning and neural network algorithms may complement each other. Our experience indicates that the neural network method requires less analyst intervention and is easier to run than is the evidential reasoning algorithm. We also found that the neural network generated results more compatible with the geological map made manually. Therefore, we conclude that neural networks are more adaptive to noise and data variability.

References

An, P., W.M. Moon, and G.F. Bonham-Carter, 1992. On knowledge-based approach of integrating remote sensing, geophysical and geological information, *IGARSS'92*, Houston, Texas, May 1992, pp. 34–38.

Azimi–Sadjadi, M.R., S. Chaloum, and R. Zoughi, 1993. Terrain classification in SAR images using principal components analysis and neural networks, *IEEE Trans. on Geos. and Remote Sens.*, 31(2):511–515.

Benediktsson, J.A., P.H. Swain, and O.K. Esroy, 1993. Conjugate-gradient neural networks in classification of multisource and very-high-dimensional remote sensing data, *Int. J. of Remote Sensing*, 14(15):2883–2903.

Bezdek, J.C., R. Ehrlich, and W. Full, 1984. FCM: the fuzzy c-means clustering algorithm, *Computers and Geoscience*, 10:191–203.

Chen, J., P. Gong, J. Nie, and J.A.R. Blais, 1993. Application of neural networks in forest ecological classification, *Technical Papers of ASPRS/ACSM Annual Convention*, New Orleans, Louisiana, 3:65–71.

Chung, C.F., P. Gong, A.N. Rencz, and M. Schau, 1993. Geological mapping in Melville Peninsula, Northwest Territories, Canada using multisource remote sensing and geophysical data, *IGARSS'93*, Tokyo, Japan, 18–21 August, pp. 913–916.

Chung, C.F., and W.M. Moon, 1991. Combination rules of spatial geoscience data for mineral exploration, *Geoinformatics*, 2:159–169.

Cibula, W.G., and H.M. Nyquist, 1987. Use of topographic and climatological models in geographical data base to improve Landsat MSS classification for Olympic National Park, *Photogrammetric Engineering & Remote Sensing*, 53(1):67–76.

Civco, D.L., 1993. Artificial neural networks for land-cover classification and mapping, *Int. J. Geographical Information Systems*, 7(2):173–186.

Cohen, J., 1960. A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, 20(1):37–46.

Corr, D.G., A.M. Tailor, A. Cross, D.C. Hogg, D.H. Lawrence, D.C. Mason, and M. Petrou, 1989. Progress in automatic analysis of multi-temporal remotely-sensed data, *Int. J. of Remote Sensing*, 10(7):1175–1195.

Dreyer, P., 1993. Classification of land cover using optimized neural nets on SPOT data, *Photogrammetric Engineering & Remote Sensing*, 59(5):617–621.

Duda, R.O., and P.E. Hart, 1973. *Pattern Classification and Scene Analysis*, Wiley and Sons, New York, 482 p.

Durrand, J.M., and Y.H. Kerr, 1989. An improved decorrelation method for the efficient display of multispectral data, *IEEE Trans. on Geos. and Remote Sens.*, 27(5):611–619.

Fleiss, J.L., J. Cohen, and B.S. Everitt, 1969. Large sample standard

errors of Kappa and weighted Kappa, *Psychological Bulletin,* 72(5):323–327.

Freeman, J.A., and D.M. Skapura, 1991. *Neural Networks: Algorithms, Applications and Programming Techniques,* Addison-Wesley, New York.

Goldberg, M., D.G. Goodenough, M. Alvo, and G.M. Karam, 1985. A hierarchical expert system for updating forestry maps with Landsat data, *Proceedings of IEEE,* 73(6):1054–1063.

Gong, P., 1994. Integrated analysis of spatial data from multiple sources, an overview, *Canadian J. of Remote Sens.,* 20(4):349–359.

Gong, P., A. Zhang, J. Chen, R.J. Hall, I.W.G. Corns, 1994. Ecological land systems classification using multisource data and neural networks, *GIS'94,* Vancouver, B.C., February, pp. 659–664.

Gong, P., and D.J. Dunlop, 1991. Comments on Skidmore and Turner's supervised nonparametric classifier, *Photogrammetric Engineering & Remote Sensing,* 57(1):1311–1313.

Gordon, J., and E.H. Shortliffe, 1985. A method for managing evidential reasoning in a hierarchical hypothesis space, *Artificial Intelligence,* 26:323–377.

Hepner, G.F., T. Logan, N. Ritter, and N. Bryant, 1990. Artificial neural network classification using minimal training set, *Photogrammetric Engineering & Remote Sensing,* 56(4):469.

Hutchinson, C.F., 1982. Techniques for combining Landsat and ancillary data for digital classification improvement, *Photogrammetric Engineering & Remote Sensing,* 48(1):123–130.

Jensen, J.R., 1986. *Introductory Digital Image Processing: A Remote Sensing Perspective,* Prentice-Hall, Englewood Cliffs, New Jersey.

Kontoes, C., G.G. Wilkinson, A. Burrill, S. Goffredo, and J. Megier, 1993. An experimental system for the integration of GIS data in knowledge-based image analysis for remote sensing of agriculture, *Int. J. Geographical Information Systems,* 7(3):247–262.

Kruse, R., E. Schwecke, J. Heinsohn, 1991. *Uncertainty and Vagueness in Knowledge Based Systems,* Springer-Verlag, New York.

Lee, T., J.A. Richards, and P.H. Swain, 1987. Probabilistic and evidential approaches for multisource data analysis, *IEEE Trans. on Geos. and Remote Sens.,* 25(3):283–293.

Liu, C., and P. Burrough, 1987. Fuzzy reasoning: a new quantitative aid for land evaluation, *Soil Survey and Land Evaluation,* 17(2):69–80.

Marble, D.F., and D.J. Peuquet, 1983. Geographic information systems and remote sensing, *Manual of Remote Sensing,* Second Edition (R.N. Colwell, editor), American Society for Photogrammetry and Remote Sensing, Vol. 1, pp. 923–958.

Moon, W.M., 1990. Integration of geophysical and geological data using evidential belief function, *IEEE Trans. on Geos. and Remote Sens.,* 28(4):711–720.

Mulder, J., and I. Corns, 1993. A decision support system for predicting and consolidating ecosystems from existing map data, *GIS'93,* Vancouver, B.C., pp. 153–159.

NCGIA, 1989. The research plan of the National Center for Geographical Information and Analysis, *Int. J. Geographical Information Systems,* 3(2):117–136.

Pao, Y., 1989. *Adaptive Pattern Recognition and Neural Networks,* Addison and Wesley, New York.

Peddle, D., 1993. An empirical comparison of evidential reasoning, linear discriminant analysis and maximum likelihood algorithms for alpine land cover classification, *Canadian J. of Remote Sens.,* 19(1):31–43.

Peddle, D.R., G.M. Foody, A. Zhang, S.E. Franklin, and E.F. LeDrew, 1994. Multisource image classification II: An empirical comparison of evidential reasoning and neural network approaches, *Canadian J. of Remote Sens.,* 20(4):396–407.

Richards, J.A., 1986. *Remote Sensing Digital Image Analysis: An Introduction,* Springer-Verlag, New York.

Robinson, A.H., R.D. Sale, and J.L. Morrison, 1984. *Elements of Cartography,* John Wiley & Sons, New York.

Rumelhart, D.E., G.E. Hinton, and R.J. Williams, 1986. Learning internal representations by error propagation, *Parallel Distributed Processing – Explorations in the Microstructure of Cognition,* Vol. 1 (D.E. Rumelhart and J.L. McClelland, editors), The MIT Press, Massachusetts, pp. 318–362.

Salu, Y., and J. Tilton, 1993. Classification of multispectral image data by binary diamond neural network and by nonparametric, pixel-by-pixel methods, *IEEE Trans. on Geos. and Remote Sens.,* 31(3):606–616.

Shafer, G., 1976. *A Mathematical Theory of Evidence,* Princeton University Press, Princeton.

Shafer, G., and R. Logan, 1987. Implementing Dempster's rule for hierarchical evidence, *Artificial Intelligence,* 33:271–298.

Shi, G., 1994. *Evidential Reasoning for Geological Mapping with Multisource Spatial Data,* Masters Thesis, Department of Geomatics Engineering, The University of Calgary, Alberta, Canada.

Skidmore, A.K., and B.J. Turner, 1988. Forest mapping accuracies are improved using a supervised nonparametric classifier with SPOT data, *Photogrammetric Engineering & Remote Sensing,* 54(10):1415–1421.

Srinivasan, A., and J.A. Richards, 1990. Knowledge-based techniques for multi-source classification, *Int. J. of Remote Sensing,* 11(3):505–525.

Story, M., and R.G. Congalton, 1986. Accuracy assessment: a user's perspective, *Photogrammetric Engineering & Remote Sensing,* 52(3):397–399.

Sui, D., 1994. Recent applications of neural networks for spatial data handling, *Canadian J. of Remote Sens.,* 20(4):368–379.

Swain, P.H., and S.M. Davis, 1978. *Remote Sensing: The Quantitative Approach,* McGraw-Hill, New York, 396 p.

Wang, Y., and D. Civco, 1994. Evidential reasoning-based classification of multisource spatial data for improved land cover mapping, *Canadian J. of Remote Sensing,* 20(4):381–395.

Yen, J., 1989. GERTIS: A Dempster-Shafer approach to diagnosing hierarchical hypotheses, *Communications of the ACM,* 32(5):573–585.

Zadeh, 1965. Fuzzy sets, *Information and Control,* 8:338–353.

Zhu, A., and L.E. Band, 1994. A knowledge-based approach to data integration for soil mapping, *Canadian J. of Remote Sensing,* 20(4):408–418.