

Journal of
Applied Remote Sensing

**Hyperspectral image classification using
an unsupervised neuro-fuzzy system**

Caiyun Zhang
Fang Qiu

Hyperspectral image classification using an unsupervised neuro-fuzzy system

Caiyun Zhang^a and Fang Qiu^b

^aFlorida Atlantic University, Department of Geosciences, 777 Glades Road,
Boca Raton, Florida 33431
czhang3@fau.edu

^bUniversity of Texas at Dallas, Program in Geospatial Information Sciences,
800 West Campbell Road, Richardson, Texas 75080

Abstract. An unsupervised neuro-fuzzy system, Gaussian fuzzy self-organizing map (GFSOM), is proposed for hyperspectral image classification. This algorithm operates by integrating an unsupervised neural network with a Gaussian function-based fuzzy system. We also explore the potential for hyperspectral image analysis of three other artificial intelligence (AI)-based unsupervised techniques popular for multispectral image analysis: self-organizing map (SOM), fuzzy *c*-mean (FCM), and descending fuzzy learning vector quantization (DFLVQ). To apply these methods effectively and efficiently to hyperspectral imagery, an optimal learning sample selection strategy and a prototype initialization system are developed. An experimental study on classifying an EO-1/Hyperion hyperspectral image illustrates that GFSOM achieves the best accuracy, since it can model both the central tendency characteristics of input samples and capture the dispersion characteristics of data within a cluster. By adopting the system initialization approach developed here, all the AI-based techniques have the capability to classify hyperspectral images and can deliver acceptable accuracy, which could consequently accelerate their transitions from the multispectral to the hyperspectral field. © 2012 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.JRS.6.063515]

Keywords: artificial intelligence; hyperspectral imaging; remote sensing.

Paper 11217 received Oct. 10, 2011; accepted for publication Feb. 15, 2012; published online Apr. 6, 2012.

1 Introduction

One of the latest advancements in remote sensing is the advent of hyperspectral imaging spectrometers that are able to acquire data simultaneously in hundreds of bands with narrow bandwidths. Hyperspectral data can provide detailed contiguous spectral curves, a trait that traditional multispectral sensors cannot offer. For example, NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) has 224 contiguous bands with a nominal bandwidth of 10 nm for each, while the multispectral sensor Landsat thematic mapper (TM) only provides seven noncontiguous broad bands with an average bandwidth of 100 nm, which leaves large gaps between bands. This increase in spectral dimensionality not only allows the delineation of the land surface at the land use and land cover level, but also potentially enables the characterization of various minerals, soils, rocks, and vegetation at the material level.¹ However, the greater spectral detail comes at a cost of huge data volumes with high dimensionality, which poses great challenges in extracting thematic information from these images.

The conventional multispectral classification approaches, such as maximum likelihood, could not deliver satisfactory results when applied to hyperspectral images.² This is attributed to the fact that most multispectral classifiers are statistics-based, assuming multivariate normal distributions, linearity, and an absence of collinearity for the input bands. This is hardly achievable for hyperspectral images with hundreds of bands. For supervised classifiers, the number of training samples for each class, at a minimum, should be more than the number of spectral bands to avoid the creation of singular matrices.³ In practice, it is suggested that the number of training samples for each class be 10 times the number of input image bands in order to obtain acceptable results,¹ which becomes very difficult, if not completely possible, for hyperspectral

images, due to their high spectral dimensionality. For these reasons, nonparametric approaches specifically designed for processing hyperspectral data have been proposed, such as spectral angle mapper, linear spectral unmixing, and spectroscopic library matching.¹ These methods are supervised techniques based on the availability of reference endmembers, which are spectrally pure reflectance readings of different materials. Since these endmember-based approaches are nonparametric, there is no need to collect a large number of training samples for each class. However, as supervised techniques, they require *a priori* knowledge of the ground reference information in the form of a complete set of endmembers. Endmembers can be derived either from extensive *in situ* and lab work using a spectroradiometer or from a hyperspectral image if spectrally pure pixels of the materials can be identified on the image. However, *in situ* data collection to obtain ground reference information in advance is not always possible prior to image classification. Additionally, endmember determination from the image itself is very difficult, due to the existence of mixed pixels in the images. Therefore, when no or very little *a priori* ground reference information is available, unsupervised approaches appear to be an attractive alternative to the existing endmember-based methods.¹ However, most conventional unsupervised image classifiers, such as the Iterative Self-organizing Data Analysis (ISODATA) technique, were designed to analyze multispectral imagery and may be inadequate for hyperspectral data analysis for two main reasons. First, these methods are computationally intensive, because all pixels in the image must be compared with all clusters through multiple iterations in order to assign them to the closest cluster. Second, they tend to suffer from performance degradation with increased spectral dimensionality of the hyperspectral images.⁴

Artificial intelligence (AI)-based approaches, especially artificial neural network and fuzzy logic techniques, have been extensively employed to analyze multispectral images. Mas and Flores⁵ have reviewed these techniques for remote sensing applications. Like conventional image classifiers, AI-based approaches can be supervised or unsupervised. Among the various unsupervised classifiers, self-organizing maps (SOM),⁶ fuzzy *c*-means (FCM),⁷ and descending fuzzy learning vector quantization (DFLVQ)⁸ have historically been used for processing multispectral images or hyperspectral images with reduced spectral dimensionality.⁹⁻¹⁴ Despite avoiding problems with singular matrices and the need for obtaining pure pixels as with the statistics-based and endmember-based approaches, these AI-based classifiers have not been widely used in hyperspectral image classification. This may be attributed to certain limitations with these AI-based classifiers. As with the ISODATA approach, these classifiers involve computationally intensive algorithms. Further, when assigning a pixel to a specific cluster, these classifiers only consider the center of the clusters and ignore the data dispersion within each cluster.¹⁵ In addition, the exponent weighting parameter involved with some of these classifiers (such as FCM and DFLVQ) is difficult to specify.¹⁶

To address these deficiencies of the SOM, FCM, and DFLVQ approaches, we propose an unsupervised Gaussian fuzzy self-organizing map (GFSOM), a neuro-fuzzy system specifically designed for hyperspectral image classification. A neuro-fuzzy system combines the advantages of neural network and fuzzy logic systems and avoids the shortcomings of each individual system when they are used separately for image classifications. These shortcomings include the black box problem of a neural network and the fuzzy system's lack of an automatic knowledge acquisition capability.¹⁷ The GFSOM classifier is built upon the success of its supervised counterpart, Gaussian fuzzy learning vector quantization (GFLVQ).² We also investigate whether the optimal learning sample selection strategy and the prototype initialization system developed for the GFSOM system can be adopted by the SOM, FCM, and FLVQ approaches in classifying hyperspectral images.

2 Background

The basic algorithms underlying the three classic AI-based unsupervised classifiers (SOM, FCM, and DFLVQ) are described below, and their limitations are briefly examined. They will be used as benchmarks to evaluate the proposed GFSOM system.

2.1 SOM

SOM is a “winner-take-all” unsupervised learning neural network. It was first proposed by Kohonen⁶ and is often used to classify inputs into different categories. SOM has two layers: an input layer and a competitive or output layer. The input layer has the same dimensionality as the input vector, and the output layer consists of a physical net of neurons located at fixed positions. Different from many other neural networks, SOM is unique in that it is independent of any activation function, and it does not have a hidden layer. For a given input vector, outputs are computed for the neurons in the output layer, and the winner is the neuron whose weight has the minimum distance from the input vector. Then the network updates the weight of the winner and those of its predefined neighbors via a learning rule. In this way, the output neurons become selectively tuned to the input patterns presented during this competitive-learning procedure. In the case of remotely sensed image classification, for a pixel with N bands as an input vector (x_1, x_2, \dots, x_n) , the winner output neuron is decided by:

$$\min(\|x_i - c_{ij}\|)(j = 1, 2, \dots, M), \quad (1)$$

where c_{ij} is the weight linking the i 'th band input and j 'th cluster of an input image, N is the dimensionality of x (i.e., the number of bands), M is the total number of clusters (i.e., the number of output neurons), and $\| * \|$ is any norm expressing the similarity between an input vector and the prototype vector. The weights of the output neurons are then updated by

$$\Delta c_{ij} = \eta(x_i - c_{ij}), \quad \text{if output neuron } j \text{ is the winner,} \quad (2)$$

$$\Delta c_{ij} = 0, \quad \text{otherwise,} \quad (3)$$

where η is the learning rate.

Like other AI-based unsupervised algorithms, the objective of SOM learning is to generate the weights to model the prototypes, also known in the literature as the code vectors or center of the clusters.¹⁶ SOM uses the Euclidean distance to determine the closeness of an input pixel to the center of a cluster, but it does not consider the data dispersion in the cluster. SOM has been extensively applied in various areas¹⁸ but has not yet been seen in clustering hyperspectral images, due to certain limitations.¹² For example, it is difficult to select the initial neighborhood size that will be altered during the learning iterations in order to achieve “useful” results.⁸ When it is used for hyperspectral image classification, the first Kohonen heuristic rule that monotonically decreases the learning rate with time works well, but the second rule that decreases the neighborhood size adaptively may not be applicable for images with a small number of spectral clusters (say, ≤ 4). The initial values for the two parameters, η and c_{ij} , need to be set before the learning starts, and these values can greatly influence both the speed of the computations and the learning outcome.

2.2 FCM

The learning rules in SOM are basically a local updating strategy that ignores the global relationships between the winner and the rest of the weights, in that all the emphasis is put on the winning prototype.¹⁶ Unlike SOM, FCM considers the global geometric structure present in the data and updates all the prototypes during the learning iterations. FCM was firstly developed by Dunn¹⁹ and improved by Bezdek.⁷ It has been widely utilized in pattern recognition, but its application in hyperspectral image classification is rare. The algorithm for FCM is based on minimization of an objective function:

$$J_m = \sum_{i=1}^N \sum_{j=1}^M u_{ij}^m \|x_i - c_{ij}\|^2, \quad 1 \leq m < \infty, \quad (4)$$

where m is the exponential weight that controls the degree of fuzziness of the clusters (and can be set as any real number greater than 1), x_i and c_{ij} are the same parameters as in Eq. (1), u_{ij} is the

fuzzy membership of x_i in the j 'th cluster, N is the dimensionality of x , M is the total number of clusters, the same as the parameters in SOM, and $\| * \|$ is any norm expressing the similarity between an input vector and the prototype vector. The process of fuzzy clustering is carried out via an iterative optimization for the objective function [Eq. (4)]. FCM updates membership u_{ij} and prototype c_{ij} by

$$u_{ij} = \left(\sum_{k=1}^M \left(\frac{\|x_i - c_{ij}\|}{\|x_i - c_{ik}\|} \right)^{\frac{2}{m-1}} \right)^{-1}. \quad (5)$$

$$c_{ij} = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m}. \quad (6)$$

FCM requires setting an appropriate value for m , an important parameter depicting the fuzziness of the system. If m is small (e.g., close to 1), FCM tends to produce almost crisp vectors, and only the winner prototype is updated, leading to a process similar to SOM. If m is big (e.g., >7), FCM becomes fuzzy and updates every prototype to a very small extent. To achieve a satisfactory result, neither very small nor very big values of m are desirable.¹⁶ Similarly, c_{ij} or u_{ij} needs to be initialized, because they determine the speed of convergence and its outcome. FCM does not explicitly consider the data dispersion of each cluster, although it does model the global geometric structure of the data. For FCM, the membership u_{ij} needs to be calculated for all the learning samples in each run, which may cause tremendous computation times when used to process hyperspectral data.

2.3 DFLVQ

DFLVQ was developed by Tsao et al.⁸ to improve the performance of SOM and FCM. It was proposed to combine the hard Kohonen's SOM and the soft FCM algorithms. In DFLVQ, the learning rate η in SOM is renamed as α and is calculated as a function of the fuzzy membership u_{ij} in FCM as

$$\alpha_{ij,t} = (u_{ij,t})^{m_t}, \quad (7)$$

where $u_{ij,t}$ and m_t are the same parameters as u_{ij} and m , respectively, as defined in the FCM algorithm [see Eq. (5)]. The subscript t refers to the current accumulated learning cycle. The updating strategy for the weights is revised as:

$$c_{ij,t} = c_{ij,t-1} + \frac{\sum_{i=1}^N \alpha_{ij,t} (x_i - c_{ij,t-1})}{\sum_{i=1}^N \alpha_{ij,t}}. \quad (8)$$

We can see that in DFLVQ, the learning rule depends not only on the distance between the input spectral vector and the prototype, but also the learning rate that is adjusted over time with m_t . A larger m_t leads to a lower learning rate, whereas a smaller m_t results in a higher learning rate. Thus, it is preferable to vary m_t in order to control the amount of fuzziness by adjusting it from a larger initial value (m_0) to a smaller final one (m_f) using a simple linear function decreasing with time. Hence, this method is referred to as DFLVQ in the literature. This improved algorithm not only adjusts the learning rate but also updates all neighborhoods with time, thus achieving a good balance between SOM and FCM. However, it is often criticized because of the difficulty in setting the right range for m and the appropriate initial values for c_{ij} , which may severely influence the computation cost and results. Again, this method does not consider data dispersion within each cluster explicitly, and it is computationally very intensive. Consequently, practical applications of DFLVQ in hyperspectral image classification are very limited.

3 GFSOM

As mentioned above, SOM, FCM, and DFLVQ only model the center of a cluster without explicitly considering the dispersion of data within the cluster in the learning procedure. To address this problem, we propose a neuro-fuzzy system that utilizes the Gaussian membership function to fuzzify the SOM neural network, which leads to a fuzzy GFSOM system. The development of this unsupervised GFSOM was inspired by its counterpart, the supervised fuzzy learning vector quantization GFLVQ,² which was first developed by Qiu and Jensen¹⁷ in an attempt to open the black box of neural systems in multispectral image classification.

The structure of the GFSOM system is similar to that of Kohonen's SOM. It has two layers: the input layer and the competitive layer, also known as the output layer (Fig. 1). The neurons in the input layer correspond to all the hyperspectral bands of an input pixel, and the number of output neurons is associated with the number of resulting clusters defined by users. In addition to these fundamental components of the SOM neural network, the GFSOM system has four further components: system initialization, fuzzification, neuro-fuzzy learning, and clustering and defuzzification.

3.1 System Initialization

3.1.1 Selection of learning samples

Unsupervised classification searches for natural groupings of spectral properties of input pixels. The selection of the input pixels used as learning samples is critical for the performance of unsupervised AI-based approaches to hyperspectral image classification. For multispectral image classifications, the unsupervised methods often use all the pixels of an image scene as the learning samples. This strategy may be feasible for small hyperspectral images but is unlikely to succeed when processing regular hyperspectral images with hundreds of bands. Even if it succeeds, the learning procedure is likely to be extremely time-consuming for these types of images. The use of all the pixels as learning samples may also encounter problems due to spatial autocorrelation, which can lead to an overestimation of the contrast between categories.²⁰ Tubbs and Coverly²¹ suggest that classification algorithms need to be modified in order to take into account the problems caused by the spatial autocorrelation structure of homogenous samples. To avoid such problems, Craig²² recommends sample pixels no closer than every 10th pixel for accuracy assessment. Therefore, using all pixels as learning samples is not only inefficient but also theoretically problematic for clustering hyperspectral images.

To address this problem, we propose a random sampling scheme (RSS) to select learning samples during the learning iteration for GFSOM, a method that can also be used by the other

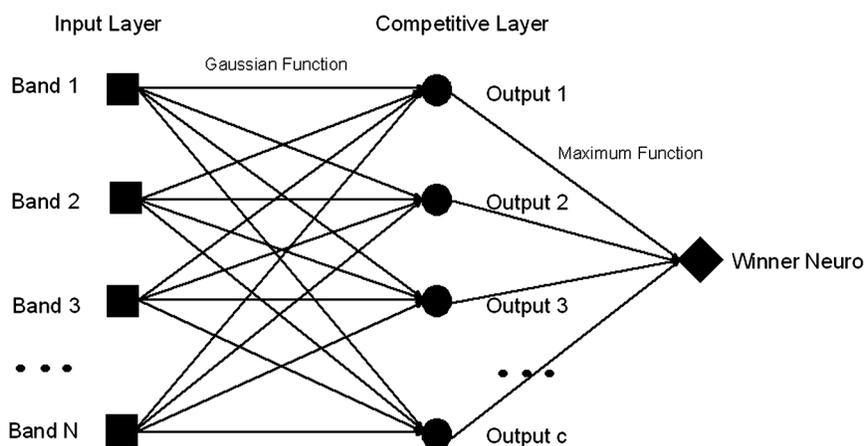


Fig. 1 The topological structure of the GFSOM system.

three AI-based unsupervised classifiers for hyperspectral images. With the RSS method, an input image is randomly sampled based on a user-defined number. This number, representing the number of pixels to be randomly selected, is set depending on the number of output clusters, the size of an input image, and the degree of spatial autocorrelation existing among the input pixels. One case study illustrated that a hyperspectral image with 160,000 pixels and 112 spectral bands only needs approximately 1000 randomly selected learning pixels to generate eight stable clusters. In each iterative learning step, the randomly selected pixels are used to update the weights of the AI-based classifiers. This proposed RSS method can greatly improve the efficiency of the learning process and reduce the spatial autocorrelation among the learning data.

3.1.2 Initialization of the prototypes

As with SOM, FCM, and DFLVQ, prototypes (c_{ij}) need to be initialized before the self-learning starts. Random assignment of prototypes is often used by the existing AI-based approaches to multispectral image classification. In SOM and GFSOM, a winner will be determined for each input learning sample. When the random assignment of prototype is used, the first winner will always be the winner, because the spectral profiles of other prototypes are still random. This leads to the learning procedure completely failing. For FCM and DFLVQ, starting from random spectral profiles of the prototypes may cause the learning procedure to be extremely slow. Thus, we propose a scheme for generating initial prototypes from the selected learning samples. This initialization strategy is based on a simplified K-mean scheme (SKS) and involves several steps. The first step is to randomly select initial prototypes from the first M learning samples. The second step is to classify the remaining learning samples to the closest prototype to form clusters. The last step is to calculate the new prototypes based on the samples assigned to individual clusters. The SKS is different from K-mean clustering, because it needs neither to cluster all of the pixels of an image nor to iterate over multiple passes. The prototypes derived from this SKS approach can then be used as the initial prototypes for GFSOM. Similarly, this scheme can also be adopted by SOM, FCM, and DFLVQ to classify hyperspectral images.

In GFSOM, it is also necessary to initialize the standard deviations that measure the dispersion of data within a cluster, a parameter lacking from the other three AI-based unsupervised classifiers. The initial value for this parameter can be calculated by applying the normal standard deviation equation to the initial value of a prototype (as the mean) and the values of the selected learning samples assigned to its related cluster. Finally, prior to running the learning procedure, all the pixel values of the selected learning samples and the original input image need to be normalized to $[0, 1]$ as in the SOM neural network.

3.2 Fuzzification

Since the input pixel values fed into the GFSOM system are not fuzzy numbers, they must be converted into a set of fuzzy numbers through a fuzzification process. The fuzzification of a pixel value of a single band is based on the Gaussian fuzzy membership function:

$$u_{ij} = e^{-\frac{(x_i - c_{ij})^2}{2 \times \sigma_{ij}^2}}, \quad (9)$$

where u_{ij} is the fuzzy membership grade, c_{ij} is the mean parameter of the Gaussian function, corresponding to the center of the j 'th cluster of the i 'th band, x_i is the i 'th input variable (i.e., the pixel value for the i 'th band) of the input learning pixel, and σ_{ij} represents the Gaussian standard deviation parameter characterizing dispersion of data in the cluster. The parameters u_{ij} , c_{ij} , x_i , i , and j are the same parameters as in FCM and DFLVQ. The GFSOM system uses the derived fuzzy membership grade, rather than the Euclidean distance, as a relative distance to determine the closeness of the input pixel to a cluster prototype. This fuzzification process models the spectral properties of input data not only by capturing their center tendency characteristics but also by modeling their data dispersion distinctiveness. The values c_{ij} and σ_{ij} are initialized using the proposed prototype initialization system. Only the selected learning samples are entered into the learning iterations.

Equation (9) can only be used to determine the fuzzy membership grade of one band for an input pixel. For hyperspectral data, to assign a pixel to a particular cluster, all the bands of this pixel need to be considered. To achieve this, an *and-or* fuzzy operator in the form of the geometric mean is utilized, so that an overall membership grade for this pixel can be obtained as:

$$\alpha_j = \left\{ \prod_{i=1}^N e^{-\left[\frac{(x_i - c_{ij})^2}{2\sigma_{ij}}\right]} \right\}^{\frac{1}{N}}, \quad (10)$$

where α_j refers to the overall membership grade of the input pixel concerning the j 'th output cluster, N refers to the number of input neurons (i.e., image bands), and c_{ij} , x_i , i , j , and σ_{ij} are the same as in Eq. (9). This is referred to as the *and-or* fuzzy operator, because it sits between the fuzzy *and* (intersection) and the fuzzy *or* (union) operators. As an averaging operator, it allows a low membership grade in one band to be compensated for by a high membership grade in another band, so that a missing or noisy value in one band will not heavily affect the clustering output of the entire pixel. In addition, the *and-or* operator is an "idempotent" function, such that *and-or* (a, a, \dots, a) = a , which delivers a more reasonable final membership grade. The winner is then assigned to the output neuron with the maximum overall membership grade.

3.3 Neuro-Fuzzy Learning

GFSOM performs fuzzy partitioning of the input vector based on competitive learning as in SOM, but unlike the original SOM that searches only the center of each cluster, GFSOM seeks to learn both the center of a cluster and the average deviation from the center of this cluster based on the pixels assigned to it. Thus, the parameters c_{ij} and σ_{ij} in the Gaussian membership function are used in conjunction to determine the "winner" output neuron. The GFSOM updates both parameters for the output neuron. Unlike the supervised system, where these two parameters are updated based on the true target class information provided by the training data and the competition result, the unsupervised learning of GFSOM has to rely solely on the competition outcome. The fuzzy competitive learning for these two parameters in GFSOM is as follows:

$$\Delta c_{ij} = \eta(x_i - c_{ij}), \quad \text{if the } j\text{'th output neuron is the winner,} \quad (11)$$

$$\Delta c_{ij} = 0, \quad \text{otherwise.} \quad (12)$$

$$\Delta \sigma_{ij} = \eta(|x_i - c_{ij}| - \sigma_{ij}), \quad \text{if the } j\text{'th output neuron is the winner,} \quad (13)$$

$$\Delta \sigma_{ij} = 0, \quad \text{otherwise.} \quad (14)$$

The updating rule for c_{ij} is the same as that used in SOM [see Eqs. (2) and (3)]. However, additional updating rules for the standard deviation weights were created [Eqs. (13) and (14)]. When the absolute deviation of the input pattern x_i from the center of the matched cluster (centered at c_{ij}) is larger than the current standard deviation σ_{ij} of the cluster, the standard deviation weight will be increased by η portion of the difference. If the absolute deviation is smaller, then the standard deviation will be decreased by that small portion of the difference. No update is needed if they are the same. This ensures that the size of the cluster will be shrunk or enlarged adaptively based on the deviations of the matched input patterns from the cluster center. The updating rule for the standard deviation works well both in the supervised GFLVQ and unsupervised GFSOM systems.

3.4 Clustering and Defuzzification

When the centers (c_{ij}) and standard deviations (σ_{ij}) of all clusters have been fine tuned in the learning procedure, they can then be directly used to cluster all the pixels of the input image.

For a hard clustering algorithm, the result is a single clustering map where each pixel is assigned to only one cluster. The output of the GFSOM system, however, is fuzzy membership grades for all clusters, affording much richer information that can be used to estimate various constituents found in a mixed pixel.¹ If a hard clustering map is desired, defuzzification, a reverse of the fuzzification process, can be used to convert various fuzzy membership grades to a single cluster for each pixel. Defuzzification is achieved by comparing a pixel's membership grades for all clusters and then assigning the pixel to the cluster with the maximum membership grade. A final thematic map is obtained through visually labeling the clusters into land use/land cover classes, as with other unsupervised classification approaches.

4 Case Study

The proposed GFSOM and the other three AI-based unsupervised classifiers (SOM, FCM, and DFLVQ) were used to classify a subset of a Hyperion image acquired in Wuxi, China (centered at $+31^{\circ}33'10.80''$, $+120^{\circ}9'54.00''$) on August 19, 2004. This subset image has a total of 400×400 pixels. Hyperion is a push broom, hyperspectral imaging spectrometer on board the Earth-Observing 1 (EO-1) spacecraft. Unlike widely used airborne hyperspectral instruments, Hyperion is a satellite sensor system with a coarse spatial resolution of 30×30 m. There are 242 unique spectral channels with a bandwidth of 10 nm for each band in the visible/near infrared (VNIR) spectral region (400 to 1000 nm) and short-wave infrared (SWIR) region (900 to 2500 nm). Due to the fact that some channels have low signal strengths, and others fall into an overlap region between VNIR and SWIR, only 196 valid bands are available for use. To alleviate the collinearity problem, we selected most of the calibrated VNIR bands and discarded the overlapped SWIR bands. A uniform feature design (UFD) was then applied as suggested by Filippi and Jensen¹³ to reduce the dimensionality of the data set while retaining as much spectral shape information as possible in the SWIR range, resulting in 112 bands being used for the subsequent analysis. A VNIR color composite map of the study area is shown in Fig. 2(a) with a band combination of 993, 603, and 711 nm. Four land cover types at level I of the USGS Classification System²³ were identified in this image, including urban/built-up, agriculture, forest, and water. We also randomly collected 491 validating pixels (130 for urban, 118 for agriculture, 124 for forest, and 119 for water), which were used as references for accuracy assessment purposes.

Successful applications of AI-based classifiers in remote sensing largely depend on the setting of appropriate parameters. The common parameters that need to be set for GFSOM and the other three AI classifiers include the number of learning cycles (i.e., iterations), number of input neurons (i.e., input bands), number of output neurons (i.e., clusters), number of learning samples, and initial values of the prototypes (c_{ij}). To ensure comparability of results, these parameters should be kept consistent across the four classifiers. Several tests were conducted in order to obtain the optimal settings of these parameters. In this case study, the outputs were obtained by setting the number of learning iterations to 100, the number of input neurons equal to the number of image bands (112 in this case), the number of output neurons to eight, and the number of randomly selected learning samples in each iterative step to 1000. Aside from these common parameters, other parameters that are unique to each classifier also need to be specified. In SOM and GFSOM, the learning rate (η) was set to decrease monotonically with time from 0.5 to 0.05, following the first Kohonen heuristic rule. This range was empirically determined based on experiments. A large value for the learning rate makes the algorithm unstable, while a small value takes a long time to converge. This adaptive learning rate keeps the learning step size as large as possible, while keeping the learning procedure as stable as possible.²⁴ In FCM and DFLVQ, setting the exponential weight (m) appropriately is critical for the successful use of these two algorithms. For FCM, most users suggest a value in the range of 1.1 to 5 to yield a suitable interpretation. In this case study, a value of 3 for m generated the best results. For DFLVQ, a varying m approach was adopted with a range of 1.1 to 7.0, as suggested by Bezdek and Pal.¹⁶ All the parameters initialized for each classifier are summarized in Table 1. Tests were conducted on a PC with a 3.4-GHz Pentium CPU and 1 GB of RAM. Once the learning procedure achieves convergence, the whole image is then input into the system to

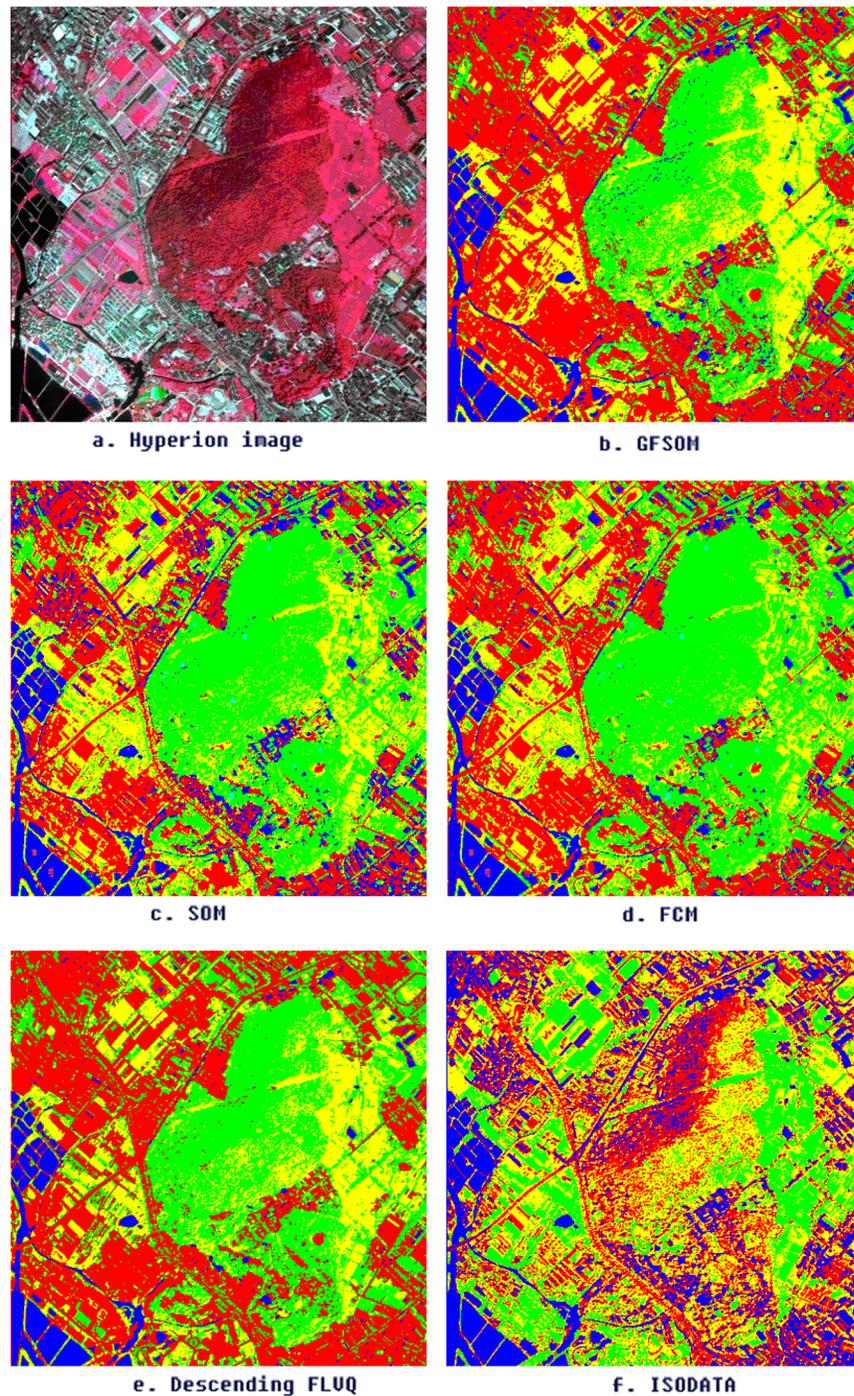


Fig. 2 (a) Hyperion image VNIR color composite map ($R = 993$ nm, $G = 603$ nm, $B = 711$ nm); (b–f) Classification results from various image classifiers (red = urban, green = forest, yellow = agriculture, blue = water).

be grouped into natural spectral clusters, which are finally transformed into classification maps with four identified classes, as presented in Figs. 2(b)–2(e).

5 Results and Discussion

To evaluate the performance of the RSS strategy for learning sample selection, classification was also attempted using all pixels (160,000) in the input image. Similarly, the random assignment

Table 1 Parameter set for SOM, FCM, DFLVQ, and GFSOM in classifying an EO-1/Hyperion hyperspectral image with 400×400 pixels and 112 bands.

Common parameters for all the classifiers	
Number of input neurons	112
Number of output clusters	8
Number of learning cycles	100
Number of selected learning samples	1000
Initialization of prototypes (c_{ij})	simplified K-mean scheme (SKM)
Parameters	SOM
Initial learning rate	0.5
Final learning rate	0.05
Parameters	FCM
Weighting exponent (m)	3
Parameters	Descending FLVQ
Initial weighting exponent (m_0)	1.1
Final weighting exponent (m_f)	7.0
Parameters	GFSOM
Initial learning rate	0.5
Final learning rate	0.05

approach to c_{ij} prototype initialization was also tested to evaluate the performance of the proposed SKS approach.

The utilization of all the pixels from the input image caused all the AI classifiers to crash before convergence, due to the large data volume of the image. This suggests that the use of all pixels of hyperspectral images as learning samples is problematic for the AI-based unsupervised classifiers. Conversely, all four classifiers were able to deliver desired results using a small number of learning samples selected with the proposed RSS strategy. We found that for the image in our case study with 160,000 pixels, we only needed to collect 1000 learning samples to achieve stable learning outcomes. Therefore, the proposed RSS strategy for learning sample selection is both an efficient and effective solution for AI-based unsupervised classification of hyperspectral images.

The SOM and GFSOM algorithms also completely failed with a random initialization approach. Although FCM and DFLVQ could converge, they took a longer time to do so. However, employing the proposed SKS approach to prototype initialization, SOM and GFSOM were able to achieve convergence, while the convergence of FCM and DFLVQ was also substantially faster. By combining the RSS learning sample selection strategy and SKS prototype initialization system, all classifiers were able to produce desired clustering maps within seconds (Table 2). SOM was the most efficient classifier and finished the self-learning process in 14 sec. GFSOM took a little longer (20 sec) to complete the self-learning process but was faster than FCM and DFLVQ, which took 35 and 33 sec, respectively.

Table 2 Computational cost of each AI-based algorithm.

	SOM	FCM	DFLVQ	GFSOM
Total CPU (sec)	14	35	33	20

Note: The computer used has a 3.4 GHz Pentium processor with 1 GB RAM.

For comparison purposes, the performance of the conventional unsupervised ISODATA method was also evaluated, and the final classification map derived from this method is shown in Fig. 2(f). The accuracy of the classifications from all the unsupervised classifiers was assessed using the confusion matrix based on 491 randomly selected samples. The producer's, user's, and overall accuracy thus obtained are given in Table 3. Kappa coefficients were also calculated for quantifying the classification accuracy. The Kappa coefficient (K_{hat}) is believed to be a better representation of the general quality of an image classification, because it removes the effects caused by differences in sample size and accounts for the off-diagonal elements in the error matrix.²⁵ It also allows different classifications to be compared statistically.²⁶ The results of these significance tests, which are based on the Z (Normal) distribution, between GFSOM and the other classifiers using the K_{hat} values are also shown in Table 3. According to Fleiss,²⁷ Kappa coefficients larger than 0.75 suggest strong agreement. Landis and Koch²⁸ suggest that Kappa coefficients larger than 0.81 indicate an almost perfect agreement. Z-scores larger than 1.96 suggest that the two methods are significantly different at the 95% statistical confidence level.

Comparing the classification results, we can see that the best overall accuracy of 88.6% was achieved by GFSOM with Kappa coefficient (K_{hat}) of 0.85. DFLVQ obtained an overall accuracy of 83.5% with Kappa coefficient of 0.78. SOM produces the poorest result among the AI-based methods with an overall accuracy of 73.1% and Kappa coefficient of 0.64, primarily caused by the misclassification of the agriculture in the east side of the image as forest and urban as water [Fig. 2(c)]. However, SOM could still produce an outcome better than ISODATA, which has an overall accuracy of only 53.4% with Kappa coefficient of 0.38 [Fig. 2(f)]. This confirms the conclusion in the literature that ISODATA is not appropriate for clustering hyperspectral data. The Z-tests indicate that the classification accuracy of GFSOM is significantly better than all other unsupervised classifiers at a 95% confidence level.

When the producer's and user's accuracy are examined, all methods generate their best accuracy for the water class and obtain varying degrees of agreement for categorizing other classes. Water is often the easiest feature to classify in remotely sensed images, due to the homogeneous nature of most water surfaces. Urban is a difficult category to identify, because of its spatial heterogeneity, resulting in larger data dispersion within this class. GFSOM achieved the best accuracy for identifying the urban class among these methods, owing to its capability to capture the dispersion of data within a class. It obtained comparable accuracy to FCM and DFLVQ for categorizing forest and agriculture classes, which have relatively homogeneous spatial patterns.

Using the proposed approaches for selecting learning samples and initializing prototypes, all AI-based classifiers were able to complete the learning and clustering process within a short period of time. As a result, the computational efficiency would not be a problem for these AI-based unsupervised classifiers in classifying hyperspectral images. In our experiment, SOM achieved the best computational efficiency, due to its simple learning process, but it had the

Table 3 Results of accuracy assessment for each algorithm.

	SOM		FCM		DFLVQ		GFSOM		ISODATA	
	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
Urban	66.2	96.6	40.7	88.9	97.7	82.5	98.5	86.5	11.5	16.5
Forest	85.4	59.9	87.9	60.6	75.8	70.7	72.6	85.7	61.9	89.0
Agriculture	40.7	84.2	93.9	89.0	66.1	83.9	83.5	84.5	44.4	44.0
Water	100	70.8	99.2	98.3	93.2	100	100	97.5	100	61.7
Overall accuracy (%)	73.1		80.9		83.5		88.6		53.4	
Kappa coefficient	0.64		0.74		0.78		0.85		0.38	
Z-score	6.38		3.45		2.33				13.33	

Note: PA, producer's accuracy (%); UA, user's accuracy (%).

lowest classification accuracy among all AI-based classifiers. As a hard classification algorithm, SOM does not have the capability to provide the rich membership information of a fuzzy classifier¹ such as FCM, DFLVQ, and GFSOM. Although FCM is a popular fuzzy classifier for multispectral data analysis, to the best of our knowledge, this study is the first to explore its applicability for hyperspectral image processing. By adopting the proposed learning sample selection and prototype initialization approaches, it has been shown to be applicable to hyperspectral data. Nevertheless, for FCM and DFLVQ, a critical issue is still the choice of an appropriate weighting exponent m , which can take any values in the range $(1, \infty)$. A varying m approach was adopted in the DFLVQ algorithm used in this case study, but its range is still difficult to determine. In contrast, GFSOM, as a fuzzy neural network relying on the Gaussian membership function adaptively learning from sampling data, does not suffer from the problem of choosing an appropriate value for parameter m to determine its fuzziness. Further, the learning algorithms of SOM, FCM, and DFLVQ all model the central tendency of the individual clusters without considering the data dispersion within each cluster. GFSOM, on the contrary, captures both the typicality of the input vectors and the atypicality of the data within a cluster using fuzzy competitive learning, which enables it to achieve the best accuracy among all the tested unsupervised classifiers. It is also the most computationally efficient classifier, since it only updates the winners in the learning iterations.

6 Summary

We proposed an unsupervised GFSOM classifier for hyperspectral image analysis with the purpose of solving the various problems inherent in the three standard AI-based approaches: SOM, FCM, and DFLVQ. These three methods have been primarily employed for multispectral image analysis with limited hyperspectral application. We thus additionally explored the potential of these techniques for hyperspectral image processing. To apply these AI-based methods effectively and efficiently, we developed a learning sample selection strategy and prototype initialization process. A case study in classifying an EO-1/Hyperion image illustrated that the proposed GFSOM achieved the best classification result among the unsupervised classifiers tested, owing to its ability to model explicitly not only the center tendency of data groups, but also the data dispersion characteristics within the groups. The results also demonstrate that all the AI-based techniques have the capability to classify hyperspectral images if they adopt the learning sample selection strategy and prototype initialization process developed for GFSOM.

It is instructive to note that our proposed method achieved the best accuracy in this case study, but this does not imply the ultimate superiority of this technique over the other approaches under all circumstances. The fact that different parameters are required for different approaches does not really make them comparable in a strict sense. Similar to other classification approaches, the accuracy of GFSOM heavily depends on the number of classes to be identified, the degree of spatial heterogeneity of the image, and the associated parameters specified. The results obtained here are certainly encouraging, but further research is needed using images with more land use/cover types, as well as different spatial and spectral resolutions in order to confirm the robustness and extensibility of this method.

Acknowledgments

This research was jointly supported by the National Basic Research Program of China under Project 2009CB723903, and the National High-Tech Research and Development Program of China under Project 2008AA09A404.

References

1. J. R. Jensen, *Introductory Digital Image Processing*, 3rd Ed., Prentice Hall, Upper Saddle River, NJ (2005).
2. F. Qiu, "Neuro-fuzzy based analysis of hyperspectral imagery," *Photogramm. Eng. Rem. Sens.* **74**(10), 1235–1247 (2008).

3. J. A. Benediktsson et al., "Classification of AVIRIS data from Iceland-Geological applications," in *International Geoscience and Remote Sensing Symposium, Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation*, California Institute of Technology, Pasadena, CA, USA **IV**, pp. 2351–2353 (8–12 August 1994).
4. N. Viovy, "Automatic classification of time series (ACTS): a new clustering method for remote sensing time series," *Int. J. Rem. Sens.* **21**(6), 1537–1560 (2000), <http://dx.doi.org/10.1080/014311600210308>.
5. J. F. Mas and J. J. Flores, "The application of artificial neural networks to the analysis of remotely sensed data," *Int. J. Rem. Sens.* **29**(3), 3617–3663 (2008), <http://dx.doi.org/10.1080/01431160701352154>.
6. T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.* **43**(1), 59–69 (1982), <http://dx.doi.org/10.1007/BF00337288>.
7. J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York (1981).
8. E. C. K. Tsao, J. C. Bezdek, and N. R. Pal, "Fuzzy Kohonen clustering networks," *Pattern Recogn.* **27**(5), 757–764 (1994), [http://dx.doi.org/10.1016/0031-3203\(94\)90052-3](http://dx.doi.org/10.1016/0031-3203(94)90052-3).
9. R. L. Cannon et al., "Segmentation of a thematic mapper image using the fuzzy c-means clustering algorithm," *IEEE Trans. Geosci. Rem. Sens.* **GRS-24**, 400–408 (1986), <http://dx.doi.org/10.1109/TGRS.1986.289598>.
10. P. Blonda et al., "Application of the unsupervised fuzzy Kohonen clustering network for remote sensed data segmentation," in *Proc. of WILF '95, Italian workshop on fuzzy logic: New trends in fuzzy logic*, A. Bonarini et al., eds., World Scientific Publishing Co., River Edge, NJ, USA, pp. 143–150 (September 21–22 1995).
11. M. Ceccarelli, A. Farina, and A. Petrosino, "Fuzzy unsupervised terrain classification based on a multiresolution approach," in *Proc. of WILF'95, Italian workshop on fuzzy logic: New trends in fuzzy logic*, A. Bonarini, D. Mancini, F. Masulli, and A. Petrosino, eds., River Edge, NJ World Scientific, Naples, Italy, pp. 151–159 (September 21–22 1995).
12. A. Baraldi et al., "Model transitions in descending FLVQ," *IEEE Trans. Neural Network* **9**(5), 724–738 (1998), <http://dx.doi.org/10.1109/72.712148>.
13. A. M. Filippi and J. R. Jensen, "Fuzzy learning vector quantization for hyperspectral coastal vegetation classification," *Rem. Sens. Environ.* **100**(4), 512–530 (2006), <http://dx.doi.org/10.1016/j.rse.2005.11.007>.
14. A. M. Filippi et al., "Self-organizing map-based applications in remote sensing," in *Self-Organizing Maps*, G. K. Matsopoulos, ed., <http://sciyo.com/articles/show/title/self-organizing-map-based-applications-in-remote-sensing> (4 December 2010).
15. T. Nomura and T. Miyoshi, "An adaptive rule extraction with the fuzzy self-organizing map: a comparison with other methods," in *Proc. of the 3rd International Symposium on Uncertainty Modeling and Analysis and Annual Conf. of the North American Fuzzy Information Processing Society*, IEEE Computer Society, College Park, MD, pp. 311–316 (September 17–20 1995).
16. J. C. Bezdek and N. R. Pal, "Two soft relatives of learning vector quantization," *Neural Netw.* **8**(5), 729–743 (1995), [http://dx.doi.org/10.1016/0893-6080\(95\)00024-T](http://dx.doi.org/10.1016/0893-6080(95)00024-T).
17. F. Qiu and J. R. Jensen, "Opening the black box of neural networks for remote sensing image classification," *Int. J. Rem. Sens.* **25**(9), 1749–1768 (2004), <http://dx.doi.org/10.1080/01431160310001618798>.
18. T. Villmann and E. Merényi, "Extensions and modifications of the Kohonen-SOM and applications in remote sensing image analysis," in *Self-Organizing Maps: Recent Advances and Applications*, U. Seiffert and L. C. Jain, eds., Springer, Berlin, pp. 121–145 (2001).
19. J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.* **3**(3), 32–57 (1973).
20. J. B. Campbell, "Spatial correlation effects upon accuracy of supervised classification of land cover," *Photogramm. Eng. Rem. Sens.* **47**(3), 355–363 (1981).
21. J. D. Tubbs and W. A. Coverly, "Spatial correlation and its effect upon classification results in Landsat," in *Proc. of the 12th International Symposium on Remote Sensing of the Environment*, ISRSE, Manila, Phillipines, pp. 775–781 (1978).

22. R. G. Craig, "Autocorrelation in Landsat data," in *Proc. of the Thirteenth International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan, USA, ISRSE, pp. 1517–1524 (1979).
23. J. R. Anderson et al., *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*, USGS Professional Paper 964, Washington, D.C., 28 p. (1976).
24. J. Amini, "Optimum learning rate in back propagation neural network for classification of satellite images (IRS-1D)," *Scientia Iranica* **15**(6), 558–567 (2008).
25. R. G. Congalton, R. G. Oldwald, and R. A. Mead, "Assessing Landsat classification accuracy using discrete multivariate statistical technique," *Photogramm. Eng. Rem. Sens.* **49**(12), 1671–1678 (1983).
26. G. H. Rosenfield and K. Fitzpatrick-Lins, "A coefficient of agreement as a measure of thematic classification accuracy," *Photogramm. Eng. Rem. Sens.* **52**(2), 223–227 (1986).
27. J. L. Fleiss, *Statistical Methods for Rates and Proportions*, 2nd Ed., John Wiley & Sons, New York (1981).
28. J. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics* **33**(1), 159–174 (1977), <http://dx.doi.org/10.2307/2529310>.

Biographies and photographs of the authors are not available.