An Unsupervised Artificial Immune Classifier for Multi/Hyperspectral Remote Sensing Imagery

Yanfei Zhong, Liangpei Zhang, Bo Huang, and Pingxiang Li

***Abstract*—A new method in computational intelligence namely artificial immune systems (AIS), which draw inspiration from the vertebrate immune system, have strong capabilities of pattern recognition. Even though AIS have been successfully utilized in several fields, few applications have been reported in remote sensing. Modern commercial imaging satellites, owing to their large volume of high-resolution imagery, offer greater opportunities for automated image analysis. Hence, we propose a novel unsupervised machine-learning algorithm namely unsupervised artificial immune classifier (UAIC) to perform remote sensing image classification. In addition to their nonlinear classification properties, UAIC possesses biological properties such as clonal selection, immune network, and immune memory. The implementation of UAIC comprises two steps: initially, the first clustering centers are acquired by randomly choosing from the input remote sensing image. Then, the classification task is carried out. This assigns each pixel to the class that maximizes stimulation between the antigen and the antibody. Subsequently, based on the class, the antibody population is evolved and the memory cell pool is updated by immune algorithms until the stopping criterion is met. The classification results are evaluated by comparing with four known algorithms: K-means, ISODATA, fuzzy K-means, and self-organizing map. It is shown that UAIC is an adaptive clustering algorithm, which outperforms other algorithms in all the three experiments we carried out.**

***Index Terms*—Artificial immune system (AIS), clustering, pattern recognition, remote sensing, unsupervised classification.**

I. Introduction

VARIOUS algorithms such as maximum likelihood, parallelepiped, and minimum distance from mean have been employed in the past for classifying multi/hyperspectral data in a pixelwise manner [1]. These algorithms are based on the fact that each class of materials, in accordance to its molecular composition, has its own spectral signature. A vast majority of these are supervised algorithms, which require that the number of classes and the class distribution model be known in advance. Furthermore, these algorithms entail training samples from each class to build models for different classes.

Unsupervised classification algorithms are built to solve the site labeling problem without the need for training samples [2] (see [3] for more reasons of using unsupervised classification). For example, the familiar K-means [4] and iterative self-organizing data (ISODATA) [5] algorithms iteratively assign the pixels of an image to one of the classes. K-means finds an optimal partition of the data distribution into the requested number of subdivisions, while ISODATA is a modified version of the K-means algorithm. Both of them assign an arbitrary initial cluster vector first. The mean vectors and covariance matrix of clusters are then calculated based on the pixels in the initial cluster; pixels in the image are assigned to the closest cluster to form a new cluster and the label of each pixel is updated. The mean vectors and covariance matrix of clusters are recalculated subsequently based on the new clusters. In every iteration of the classical K-means and ISODATA algorithms, each image pixel is assumed to be in exactly one cluster, an alternative to the crisp membership association uses fuzzy sets to describe the relationship between the data points and the cluster centers. For instance, fuzzy K-means [6] is an approach to clustering those partitions of an image datasetinto K fuzzy subsets using fuzzy membership. In addition to the aforementioned algorithms, Bayesian classifiers [7] and Markov random fields [8] have also been employed to obtain relative frequencies of individual and neighbors among a pixel. Recently, there has been considerable interest in applying unsupervised neural networks [9], such as Kohonen's self-organizing maps (SOM), to multi/hyperspectral remote sensing image classification. SOM was investigated as a possible tool for automated knowledge acquisition.

Different with the above classifiers, we propose a novel unsupervised artificial immune classifier (UAIC) to perform remote sensing image classification. Artificial immune systems (AIS), which are inspired by the immune systems, use the immunological properties in order to develop adaptive systems to accomplish a wide range of tasks in various areas of research [11]–[14] including pattern recognition [15], [16], intrusion detection [17], [18], clustering [19], optimization [20], and intelligence control [21]. In spite of the successful application of AIS in several fields, few applications have been reported in remote sensing. This may be due to the fact that it is difficult to apply current AIS models to remote sensing image classification owing to the huge data volumes associated with remote sensing images. Nevertheless, we have attempted AIS for supervised remote sensing image classification [22], [23], and we will explore UAIC for unsupervised multi/hyperspectral image classification in this paper, providing that it is sometimes hard to obtain a representative set of training samples in the supervised classification.

Manuscript received March 22, 2005; revised October 11, 2005. This work was supported in part by the National Natural Science Foundation of China under Grant 40471088, in part by the 973 Program of the People's Republic of China under Grant 2003CB415205, in part by the National Science and Engineering Research Council of Canada under Grant 75-3594, and in part by the Foundation of State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing under Grant 904151695.

Y. Zhong, L. Zhang, and P. Li are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, China (e-mail: zlp62@public.wh.hb.cn).

B. Huang is with the University of Calgary, Calgary, AB T2N 1N4, Canada and also with the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China.

Digital Object Identifier 10.1109/TGRS.2005.861548

In contrast to the conventional statistical classifiers, UAIC is a self-learning algorithm by utilizing the immunological properties, such as memory property and clonal selection. The advantages of AIS can be understood from the following theoretical aspects. First and foremost, UAIC are data-driven self-adaptive methods as they adjust themselves to the data without any explicit specification of functional or distributional form for the underlying model. Second, they are universal functional approximators since UAIC can approximate any function with arbitrary accuracy. Third, UAIC are nonlinear models, and hence are flexible in modeling complex real world relationships. Last, UAIC inherits the memory property of human immune systems and can recognize the same or similar antigen quickly at different times. Our experiments elucidated that UAIC has high classification precision and can be used in remote sensing image classification.

The remainder of the paper is organized as follows. Section II provides a synopsis of the human immune system. Section III explains the UAIC in detail. In Section IV, the experimental results are provided. Section V discusses the main properties of the UAIC in theoretical and empirical terms. Finally, the conclusion is provided in Section VI.

II. Human Immune System

The human immune system is a complex system made of cells, molecules, and organs that together constitute an identification mechanism capable of perceiving and combating dysfunction from our own cells and the action of exogenous infectious microorganizms as well. The human immune system safeguards us against infectious agents such as viruses, bacteria, fungi, and other parasites. Any molecule that can be recognized by the adaptive immune system is known as an antigen (Ag). Lymphocytes or the white blood cells are the fundamental components of the immune system. Within the human body, Lymphocytes are found in two forms, B cells and T cells. Functionally, these two types of cells differ in their mode of antigen recognition. B-cells are capable of recognizing antigens free in solution, while T cells require antigens to be presented by other accessory cells. Each has its distinct chemical structure and produces many Y-shaped antibodies (Ab) from its surface to kill the antigens. Ab's are molecules attached primarily to the surface of B cells whose aim is to recognize and bind to Ags [24].

The clonal selection theory [25] explains how an immune response is mounted when a nonself-antigenic pattern is recognized by a B cell. As Fig. 1 shows, when a B-cell receptor recognizes a nonself-antigen with certain affinity, it is then selected to proliferate and produce antibodies in high volumes. The antibodies are soluble forms of the B-cell receptors that are released from the B-cell surface to cope with the invading nonself-antigen. Antibodies bind themselves to antigens, thus resulting in their eventual elimination by other immune cells. In case of immune cells, proliferation is an asexual or a mitotic process; the cells divide themselves.

Once a B cell is sufficiently stimulated through close affinity to a specific antigen, it rapidly produces clones of itself. At the same time, the B-cell clones undergo a hypermutation process at particular sites in its gene, which enables the new cells to match the antigen more closely. There is a very rapid proliferation of immune cells, successive generations of which are better and better matches for the antigens of the invading pathogen [26].

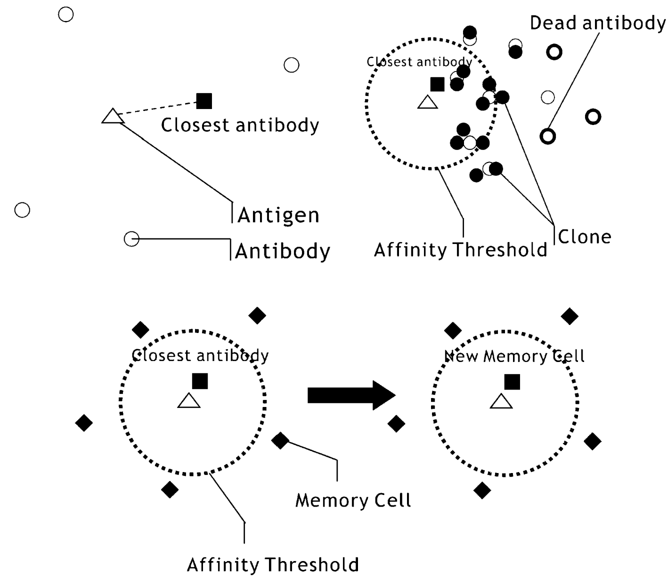


Fig. 1. Adaptive immune response. (a) Closest antibody chosen. (b) Chosen antibody clones and mutates, producing new antibody (antibody not stimulated will die). (c) Antibody with high affinity becomes memory cell.

The B cells that are not stimulated as they do not match any antigens in the body will eventually die. On the contrary, the activated B cells with high antigenic affinities are selected to become memory cells. When a body has successfully defended against a pathogen, memory cells remain and circulate in the blood, lymph, and tissues for very long periods of time. These memory cells recognize antigens similar to those that originally caused the immune response and created the memory cells, so that the body's response to a later invasion of the same pathogen or a very similar invader is much more rapid and powerful than to an invader never seen before in the primary response.

Similar to the immune systems, neural networks also create memory elements. However, their underlying mechanisms of memory and recognition are very different. In neural systems, the assimilation of memories is achieved by alteration of the strengths of connections between neurons, rather than changes within the neurons themselves. Further, the brain allows memories to be addressable by content, so that the frequent death of individual neurons does not drastically affect the performance of the brain as a whole [14], [27]. For example, in an adaptive resonance theory (ART) based network, a cluster (or class) is represented by its neural memory as a template. A template is an abstraction of the patterns in a cluster that are formed and modified during the learning (training) process. Hence, the number of templates increases (for a fixed parameter setting) according to the diversity of input patterns [28], [29].

The immune systems possess a cross-reactive memory that is observed when an individual develops a memory to one antigen and is challenged with a related, but different one. The cross-reactive memory, clonal expansion, and programmed cell death rates allow the immune system to dynamically allocate resources as needed in a distributed environment.

TABLE I
MAPPING BETWEEN THE HUMAN IMMUNE SYSTEM AND UAIC

Immune System	UAIC
Antigens	Training data/image pixel
Antibody	Feature vector/clustering centers
Shape-Space	The possible values of the data vector/pixel
Clonal Expansion	Reproduction of clustering centers that are well matched with antigens
Affinity Maturation	Proportional mutation of clustering centers and removal of lowest stimulated clustering centers
Immune Memory	Memory set of mutated clustering centers
Metadynamics	Continual removal and creation of clustering centers

III. UNSUPERVISED ARTIFICIAL IMMUNE CLASSIFIER

The clonal selection followed by the B-cells of the biological immune system is the fundamental mechanism on which UAIC is modeled. Antigens in UAIC are simulated as feature vectors which are presented to the system during training and testing. In particular, UAIC has its specific representation in remote sensing image classification. The antibodies as candidate clustering centers experience a form of clonal expansion after being presented with an input image data (analogous to antigens). When antibodies are cloned, they must undergo the affinity mutation process inversely proportional to the antigenic affinity: the higher the affinity, the smaller the mutation rate. The term metadynamics of the immune system refers to the continuous changing of the AB population through antibodies proliferation and death. The above process is described in UAIC with the continual creation and removal of antibodies with lower affinity from the population. Table I summarizes the mapping between the human immune system and UAIC.

During the course of iteration in UAIC, there may be many antibodies; however, in the final system, only their special subset constituting the memory cells will be used to classify the image in the next iteration.

Multi/hyperspectral remote sensing data $\vec{x} = \{x^1, x^2, \ldots, x^{N_b}\}^T$ through N_b bands are observed and mapped to a finite rectangular lattice $W = ((i,j)) : 1 \leq i \leq N_i, 1 \leq j \leq N_j)$, where N_i and N_j represent the number of rows and columns, respectively. The character T denotes the transpose of a matrix. The set $x^b = \{x^b_{11}, \ldots, x^b_{N_i N_j} : b = 1, \ldots, N_b\}^T$ denotes the data taken at the bth wavelength, where $x^b_{ij} \in (0, \ldots, G-1)$ and G is the number of observable gray levels. At the ij th pixel, a N_b-dimensional feature vector $x_{ij} = (x^1_{ij}, \ldots, x^{N_b}_{ij})^T$ is observed. The entire set of image data can be denoted as $x = (x_{ij} : 1 \leq i \leq N_i, 1 \leq j \leq N_j)$. A classified image is denoted as $\omega = (\omega_{ij} : 1 \leq i \leq N_i, 1 \leq j \leq N_j)$, each pixel of which is to be assigned to one of nc classes. That is, $\omega_{ij} \in (1, 2, \ldots, \mathrm{nc})$, where nc is the number of classes and is assumed to be known.

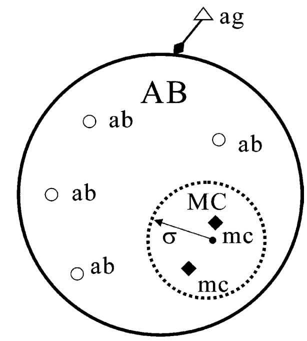


Fig. 2. Antibody population (denoted by AB) model of one class (σ represents the AB's scale/radius of influence).

Hence, for this discussion about multi/hyperspectral remote sensing classification using UAIC, let us establish the following notional conventions [30].

- Let AB represent the set of antibodies, and ab represents a single antibody where ab $\in$ AB.
- Let MC represents the set of memory cells and mc represents an individual member of this set.
- Let ag.c represents the class of a given antigen, ag, where ag.c $\in C = \{1, 2, \ldots, \mathrm{nc}\}$ and nc is the number of classes in the dataset.
- Let mc.c and ab.c represent the class of a given memory cell and antibody, mc and ab, respectively, where mc.c $\in C$, ab.c $\in C$.
- Let MC_c represent the memory cell's set of the cth class such that $\mathrm{MC}_c \subseteq \mathrm{MC} = \{\mathrm{MC}_1 \cup \mathrm{MC}_2 \cup \cdots \mathrm{MC}_{\mathrm{nc}}\}$ and $\mathrm{mc} \in \mathrm{MC}_c = \{\mathrm{mc} | \mathrm{mc.c} \equiv c\}$.
- Let AB_c represent the antibody's set of the cth class such that $\mathrm{AB}_c \subseteq \mathrm{AB} = \{\mathrm{AB}_1 \cup \mathrm{AB}_2 \cup \cdots \mathrm{AB}_{\mathrm{nc}}\}$ and $\mathrm{ab} \in \mathrm{AB}_c = \{\mathrm{ab} | \mathrm{ab.c} \equiv c\}$.
- Let ag.f and mc.f represent the feature vector of a given antigen and memory cell, ag and mc, respectively. Let ag.f_i represent the value of the ith value of ag.f and mc.f_i represents the value of the ith value of mc.f. In remote sensing image classification, they represent the gray value of every band.

With the above notations, the AB model can then be built. Fig. 2 shows a diagrammatic representation of the notion of Antibody (AB) set model of one class: there is a certain volume AB in the immune system that contains many antibodies of the class (represented by the circles and denoted by ab) and memory cells (represented by the rhombi and denoted by mc). In AB, there is a small surrounding region called memory cell set contained all memory cells of the class, denoted by MC. In remote sensing image classification, the memory set decides the recognizing ability of the whole AB. In Fig. 2, σ represents the MC's scale/radius of influence. Within the range of σ, the AB can recognize all antigens. That is, the AB can represent a number of antigens. As can be seen from Figs. 1 and 2, upon encountering an antigen, antibodies (ab in Fig. 2) are stimulated undergoing cloning and mutation. The antigens are then attacked by antibodies and removed from the immune systems. The immune systems maintain and evolve the memory set (MC in Fig. 2) so that if ever exposed to the same antigen a quicker response can be elicited against the infection. he proposed algorithm is as follows.

Step 1. Select as the first memory cell the most centrally located instance

Step 2. FOR every nonselected antibody ab_i DO

Step 2.1 FOR every nonselected antibody w_i DO

Calculate $C_{ji} = \max(D_j - d_{ji}, 0)$ where $d_{ji} = dis(ab_i, ab_j)$ and

$D_j = \min_s d_{sj}$ being s one of the selected memory cells.

Step 2.2 Calculate the gain of selecting ab_i by $\sum_j C_{ji}$

Step 3.Select the not yet selected instance ab_i as memory cell mc which

maximizes $\sum_j C_{ji}$

Step 4. IF there are C selected memory cells THEN stop
ELSE go to Step 2.

Step 5.For having a clustering assign each nonselected antibody in $Ab_{\{r\}}$ to the

cluster represented by the nearest memory cell to initial the $Ab_{\{r\}}$

Fig. 3. Pseudocode of the KA initialization method.

A. Initialization

The initialization stage can be thought of as a data preprocessing stage combined with a parameter discovery stage.

UAIC applies Kaufman approach (KA) [31] to initial memory cell population MC. The algorithm is represented in Fig. 3. In this case, the initial memory cell population is obtained by the successive selection of representative instances until C memory cells have been found. The first representative memory cell is the most centrally located instance in AB. It indicates the first representative memory cell has the minimum distance to the mean of AB. The rest of representative memory cells are selected according to the heuristic rule of choosing the memory cells that promise to have around them a higher number of the rest of antibodies.

In UAIC, the function $\mathrm{dis}(x, y)$ represents the distance between vector x and y. Since UAIC is applied to multi/hyperspectral remote sensing image classification, the distance between x and y, $\mathrm{dis}(x, y)$, is calculated using the spectral angle mapping algorithm (SAM) [32]. Let vector $x = \{x^1, x^2, \ldots, x^{N_b}\}$ and $y = \{y^1, y^2, \ldots, y^{N_b}\}$, N_b is the band number of the remote sensing image. Then the distance between x and y is given by (1)

$$\mathrm{dis}(x, y) = \alpha = \cos^{-1}\left[\frac{\sum_{i=1}^{N_b} x^i y^i}{\left[\sum_{i=1}^{N_b} (x^i)^2\right]^{\frac{1}{2}} \left[\sum_{i=1}^{N_b} (y^i)^2\right]^{1/2}}\right]. \tag{1}$$

Affinity is inversely proportional to distance in the feature space. In UAIC, affinity is defined as in (2) below according to the antibody population model (Fig. 2) so that the affinity between antigens and antibodies or between two antibodies is in the range [0, 1] and each AB has its radius of influence

$$\mathrm{affinity}(x, y) = \exp\left(-\frac{\mathrm{dis}(x, y)}{2\sigma_i^2}\right) \tag{2}$$

and σ_i is AB's scale/radius of influence

B. Classification Using UAIC

Once initialization is over, the next step is the iteration of the algorithm. For each iteration, the algorithm performs the following steps to train each antigen ag in the remote sensing image.

Step 1: Assign ag to kth class: For each ag in the image, assign that antigen to one of nc classes, where the class is assumed to be the kth class. Given a specific training antigen, ag, find the memory cell, mc, that has the maximal affinity as follows:

$$\mathrm{mc} = \arg \max_{\mathrm{mc} \in \mathrm{MC}} \mathrm{affinity}(\mathrm{ag}, \mathrm{mc}). \tag{3}$$

Then assign that ag to the class of mc, $\mathrm{ag.c} \equiv \mathrm{mc.c} \equiv k (k \in C = \{1, 2, \ldots, \mathrm{nc}\})$.

Step 2: Evolving the antibody population AB^k: After assigning the ag to kth class, evolving the antibody population AB^k and the memory cell pool MC^k are accomplished as follows.

1) Determine the vector f_k that contains the affinity of ag to all the N_{AB} Ab's in AB^k, where N_{AB} is the number of the antibody set AB^k.
2) Select the n highest affinity Ab's from AB^k to compose a new set $\mathrm{AB}^k_{\{n\}}$ of high-affinity Ab's in relation to ag, where n is the number of the cloned antibodies in AB^k.
3) The n selected Ab's independently and proportional to their antigenic affinities, generating a clone set C^k: the higher the antigenic affinity, the higher the number of clones generated for each of the n selected Ab's. The number of clones generated for all these n selected antibodies is given by

$$\mathrm{NumClones} = \sum_{i=1}^{n} \mathrm{round}(\mathrm{Clonal\_rate} \bullet \mathrm{affinity}(\mathrm{ag}, \mathrm{ab}_i)) \tag{4}$$

where *NumClones* is the total number of clones generated for ag. The clonal rate, denoted by *Clonal_rate*, is used to determine how many clones are produced by Ab's and memory cells, a typical value is 10, and round() is the operator that rounds its argument toward the closest integer.
4) Submit the clones set C^k to an affinity maturation process inversely proportionally to its antigenic affinity, generating a population MU^k of matured clones: the higher the affinity, the smaller the mutation rate. The mutation rate is determined by (5) as follows:

$$\mathrm{mutate\_rate} = 1 - \mathrm{affinity}(\mathrm{ag}, \mathrm{ab}_i). \tag{5}$$

The mutation of the clones set C^k is performed according to the following equation:

$$\mathrm{MU}_i^k = C_i^k + \mathrm{mutation\_rate} \bullet N(0, 1) i \in [1, \mathrm{NumClones}] \tag{6}$$

where $N(0, 1)$ is a Gaussian random variable of zero mean and standard deviation of one. As MU_i^k represents a candidate solution, it must be within the range of the functions specified domain. If MU_i^k exceeds that, then it is rejected and removed from the population.

$CandAff = affinity(ag, mc_{candidate})$

$MatchAff = affinity(ag, mc_{match})$

$CellDis = Dis(mc_{candidate}, mc_{match})$

if(CandAff > MatchAff)

if(CellDis < DT*DTS)

$MC_k = MC_k - mc_{match}$

end

$MC_k = MC_k \cup mc_{candidate}$

end

Fig. 4. Update memory cell pool.

5) Redetermine the affinity f_k^* of the matured clones MU^k in relation to antigen ag.
6) Select the highest affinity ab in relation to ag to be a candidate memory cell, $\mathrm{mc}_{\mathrm{candidate}}$, to enter the set of memory antibodies MC^k.
7) Replace the β lowest affinity ab from AB^k with d highest affinity from MU^k in order to evolve the antibody population. β is the displace rate.

Step 3: Updating memory cell pool MC^k: The final stage in the training process is the potential introduction of the just-developed candidate memory cell, $\mathrm{mc}_{\mathrm{candidate}}$, into the set of existing memory cells MC.

1) Find the memory cell in MC^k, $\mathrm{mc}_{\mathrm{match}}$, that has the following property:

$$\mathrm{mc}_{\mathrm{match}} = \arg \max_{\mathrm{mc} \in \mathrm{MC}^k} \mathrm{affinity}(\mathrm{ag}, \mathrm{mc}). \tag{7}$$

2) Calculate the distance threshold (DT)

$$\mathrm{DT} = \sum_{i=1}^{N_b} (\mathrm{MAX}_i - \mathrm{MIN}_i) \tag{8}$$

where $\mathrm{MAX}_i, \mathrm{MIN}_i$ represent the maximum and minimum values of the ith and of the remote sensing image, respectively. N_b is the number of bands of the image.

3) Promoting candidate memory cell to memory cell pool MC^k.

The candidate memory cell is added to the set of memory cells only if it has higher affinity in relation to the training antigen, ag, than $\mathrm{mc}_{\mathrm{match}}$, where affinity is defined as in (2). If this test is cleared, then if the distance between $\mathrm{mc}_{\mathrm{candidate}}$ and $\mathrm{mc}_{\mathrm{match}}$ is less than the product of the affinity threshold and the user-defined distance threshold scalar (DTS), then $\mathrm{mc}_{\mathrm{candidate}}$ replaces $\mathrm{mc}_{\mathrm{match}}$ in the set of memory cells. The process is presented in Fig. 4.

Once the candidate memory cell has been evaluated for addition into the set of established memory cells, training on this antigen is complete. The next antigen in the multi/hyperspectral image is then selected and the training process proceeds from step 1 to step 3. This process continues until the system has been presented with all antigens in the image.

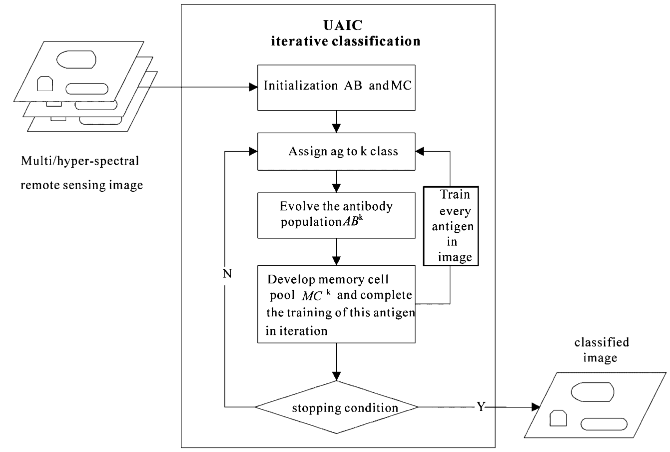


Fig. 5. UAIC flowchart.

Step 4: Consolidating and controlling the memory cell pool MC: Subsequent to each iteration, memory cells with identical session data information should be merged to limit the memory cell population growth according to their affinity.

C. Stopping Condition

The stop condition is different in different applications. One option is to set a fixed number of iterations as the stop condition. Another option is to set a fixed threshold, i.e., the pixel change threshold, for the proportion of pixels in each class that change class as the stop condition. In UAIC, the latter is selected as the stopping condition. Finally, UAIC outputs the classification result of remote sensing image. The flowchart for UAIC is shown in Fig. 5.

IV. Experiment Results

The aforementioned UAIC algorithm was coded in Visual C++6.0 and tested on different images. Three experiments were conducted to test its performance. Consistent comparisons between UAIC and traditional unsupervised algorithms, K-means and ISODATA, were completed. The estimation of classification accuracy for the several classifiers is provided.

A. Experiment 1: Wuhan TM

We tested the unsupervised classification algorithm proposed in Experiment 1 using 30-m resolution multispectral Landsat TM image shown in Fig. 6. The image (400×400 pixels), was acquired in Wuhan city, Hubei, China, On October 26, 1998. The survey area is part of the city, and the primary objective of the survey was to discriminate various objects. The observed image was expected to fall into four classes: water, vegetation, road and building. The list of classes and the number of labeled samples for each class is given in Table II.

The chief running parameters that should be provided by users in the classification calculation were as follows: the number of iterations, the number of classes, the number of highest affinity ab, clonal rate *Clonal_rate*, displace rate, and distance threshold scalar. The values of parameters were set as shown in Fig. 7. For a convenient comparison between UAIC and traditional unsupervised algorithms, the pixel

Fig. 6. Wuhan TM image, October 1998 RGB(3, 2, 1).

TABLE II
LIST OF CLASSES AND NUMBER OF LABELED SAMPLES IN EACH CLASS FOR EXPERIMENT 1

Class Name	Number of labeled samples
Water	450
Vegetation	365
Road	418
Building	429
Total number of samples	1662

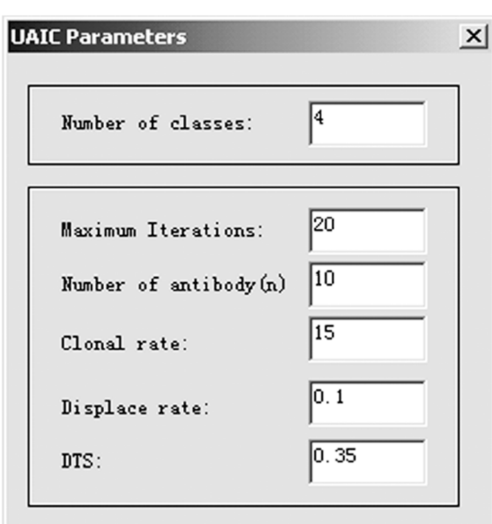


Fig. 7. Parameters dialog in Experiment 1.

change threshold as stop condition is kept at the same value, 3%. Fig. 8(a) illustrates the classification result using UAIC.

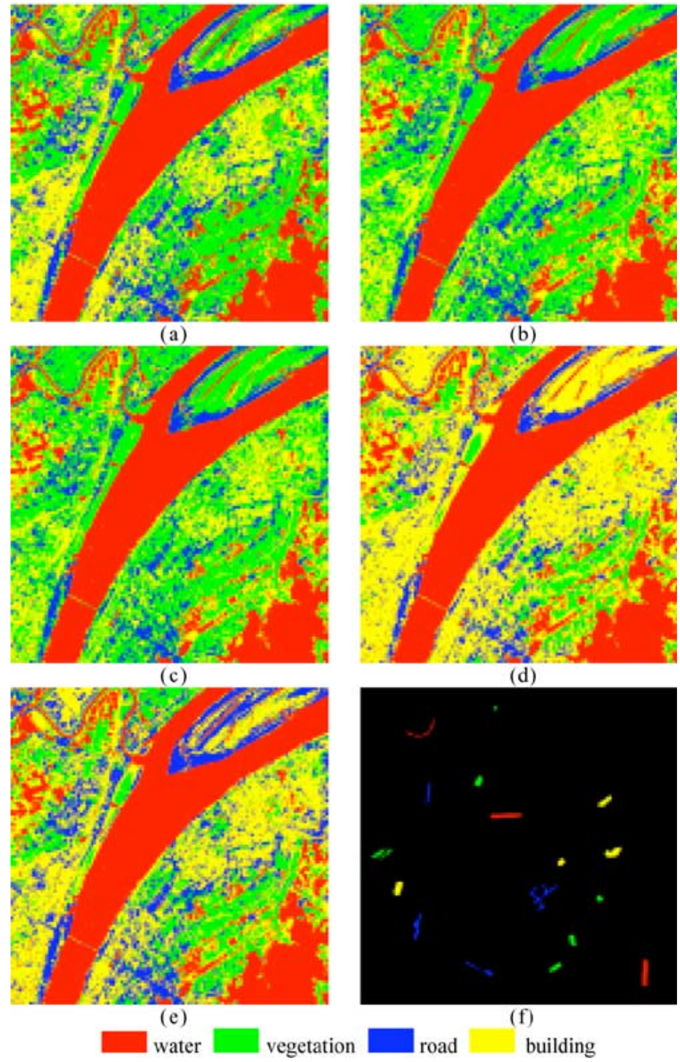


Fig. 8. Unsupervised classification images for Wuhan TM image. (a) UAIC. (b) K-means. (c) ISODATA. (d) Fuzzy K-means. (e) SOM (10×10 feature map). (f) The image for test fields used in Experiment 1.

Fig. 8(b)–(e) illustrates the classification results using K-means, ISODATA, fuzzy K-means, and SOM (using 10×10 feature map) algorithms, respectively. To evaluate the classification accuracy, a test field map is provided in Fig. 8(f) based on the ground truth data.

The visual comparisons of the five cluster classifications in Fig. 8 suggest varying degrees of accuracy of pixel assignment. It can be found from the classification images (Fig. 8) that five classifiers have similar classification results in the water class. K-means and ISODATA create similar classification maps, and it is hard to differentiate between buildings and roads. While being able to distinguish between buildings and roads, fuzzy K-means fares the worst in vegetation classification because many vegetation pixels are misclassified to the building class. SOM also recognizes the roads well though a number of non-road pixels (e.g., vegetation) are misclassified to the road class. By contrast, UAIC achieves the best visual accuracy in the vegetation class than other classifiers, and also performs satisfactorily to the building and road classes. As a result, those using UAIC have better results for four classes.

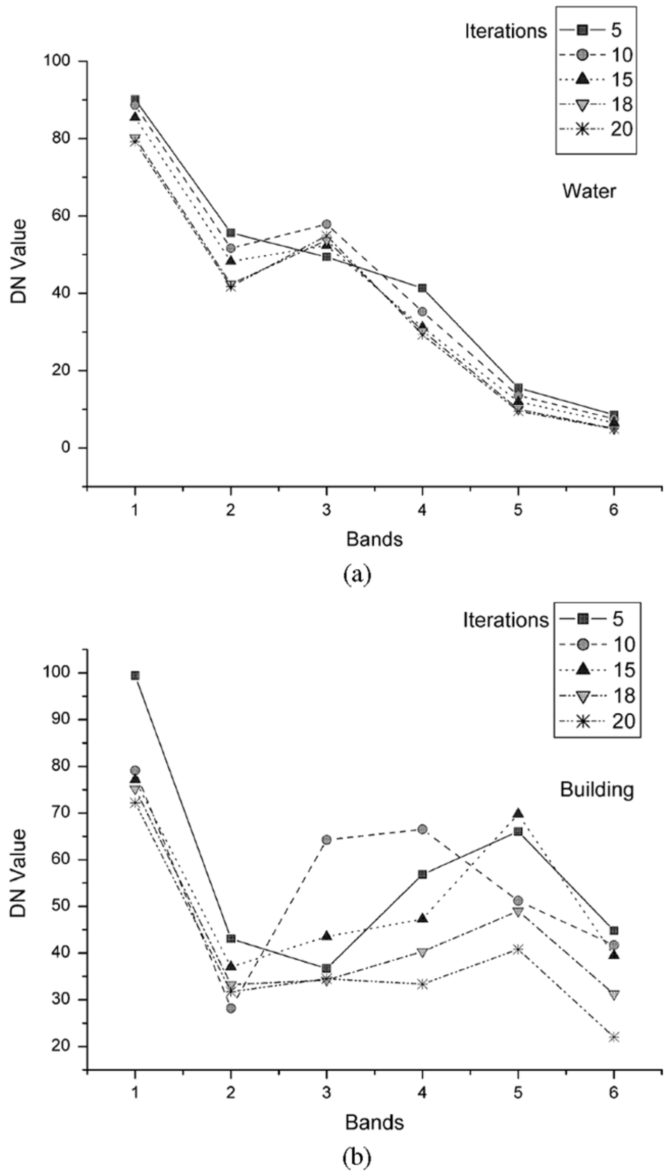


Fig. 9. (a) Convergence of memory population in the water class. (b) The convergence of memory population in the building class.

TABLE III
COMPARISON OF FIVE METHODS OF CLASSIFICATION

Methods		Water	Vegetation	Road	Building	Total
K-means	Water	**413**	22	0	0	435
	Vegetation	37	**211**	51	89	388
	Road	0	53	**295**	24	372
	Building	0	79	72	**316**	467
	Total	450	365	418	429	**1235**
ISODATA	Water	**434**	15	7	0	456
	Vegetation	16	**238**	38	91	383
	Road	0	45	**263**	21	349
	Building	0	67	90	**317**	474
	Total	450	365	418	429	**1272**
Fuzzy K-means	Water	**450**	17	5	0	472
	Vegetation	0	**219**	45	57	321
	Road	0	23	**291**	31	345
	Building	0	106	77	**341**	524
	Total	450	365	418	429	**1301**
SOM	Water	**450**	56	17	0	523
	Vegetation	0	**245**	29	24	298
	Road	0	31	**338**	88	457
	Building	0	33	34	**317**	**384**
	Total	450	365	418	429	1350
Unsupervised artificial immune classifier (UAIC)	Water	**450**	22	0	14	486
	Vegetation	0	**284**	22	22	328
	Road	0	14	**333**	59	406
	Building	0	45	63	**334**	442
	Total	450	365	418	429	**1401**

TABLE IV
COMPARISON OF FIVE CLASSIFIER PERFORMANCES IN EXPERIMENT 1

Accuracy	K-means	ISODATA	Fuzzy K-means	SOM	UAIC
Overall accuracy	74.31%	76.53%	78.28%	81.23%	84.30%
Kappa coefficient	0.6570	0.6866	0.7093	0.7486	0.7899

In executing UAIC, the population of memory cells in each class continues to change. When the mean of the population in each class is less than a fixed threshold, the iteration will end.

Fig. 9(a) and (b) shows the converging trend of the memory population by calculating the average value of memory cells in the water and building classes, respectively. As iteration goes on, the change of memory population between two successive iterations becomes smaller. It is noticed that the interval of the change in the water class is smaller than the building class. This is due to the fact that the complexity of ground substances and the diversity of disturbance are different in the two classes. In general, the water class is more homogeneous, hence the convergence in this class is faster.

For a more detailed verification of the results, we compared ground truth data with the classified images and assess the accuracy of each classifier quantitatively using both the overall accuracy measure and the Kappa coefficient. Tables III–IV list the results of comparisons between the ground truth data and classified images obtained by five classifier: UAIC, K-means, ISODATA, fuzzy K-means, and SOM.

From Tables III and IV it is apparent that the UAIC classifier produces better classification results than other classifiers. The details are as follows: UAIC exhibits the best overall classification accuracy, i.e., the best percentage of correctly classified among all the testing pixels considered, with a gain of 9.99%, 7.77%, 6.02%, and 3.07% over the K-means, ISODATA, fuzzy K-means, and SOM algorithms, respectively. UAIC improves the Kappa coefficient from 0.6570 to 0.7899, an improvement by 0.1329. This is due to that the conventional unsupervised multivariate classifiers require ideal conditions. However, because of the complexity of ground substance and the diversity of disturbance, the ideal conditions are not often met in real classification calculations. As a result, these conventional classification methods have a low precision. On the other hand, UAIC is a

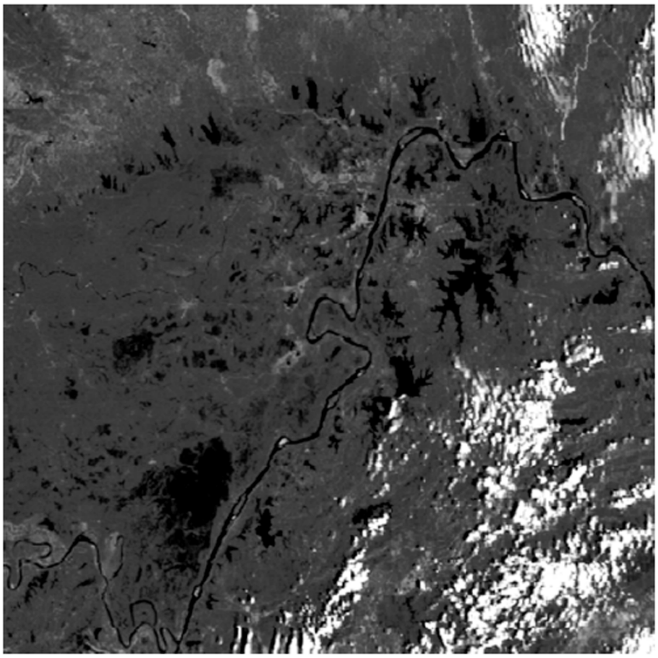


Fig. 10. Wuhan MODIS image, April 2, 2002 RGB(3, 4, 6).

TABLE V
LIST OF CLASSES AND NUMBER OF LABELED SAMPLES IN EACH CLASS FOR EXPERIMENT 2

Class Name	Number of labeled samples
Water	295
Vegetation	263
City	298
Cloud	263
Total number of samples	1119

data-driven self-adaptive method which can adjust itself to the data without any explicit specification of functional or distributional form for the underlying model. UAIC can approximate any function with arbitrary accuracy by universal functional approximator. In addition, UAIC is a nonlinear model which makes it flexible in modeling real, complex relationships. Therefore, UAIC classifier has the capacities of self-learning and is robust.

B. *Experiment 2: Wuhan MODIS*

Wuhan is the study area considered in this experiment. The data item employed in this experiment is a 500-m resolution MODIS image (400×400 pixels), acquired on April 2, 2002. The level 1B datasets include 500-m reflectance data for channels 3, 4, 6, 7. The four spectral channels used in this experiment are at 0.46–0.48, 0.55–0.57, 1.63–1.65, and 2.11–2.16 μm. Fig. 10 shows the experimental MODIS image. The observed image was expected to fall into four classes: water, vegetation, city, and cloud. The list of classes and the number of labeled samples for each class is given in Table V.

In Experiment 2, the parameters were set as shown in Fig. 11. It should be noted that UAIC has a different value of DTS in Experiment 2, from 0.35 in Experiment 1 to 0.7 in Experiment 2. For a convenient comparison between UAIC and traditional unsupervised algorithms, the pixel change threshold is kept at the same value, 3%. Fig. 12(a) illustrates the classification result using UAIC. Fig. 12(b)–(e) illustrates the classification results using K-means, ISODATA, fuzzy K-means, and SOM algorithms. To evaluate the classification accuracy, a test field map is provided in Fig. 12(f) based on the ground truth data. The classification accuracy for the several classifiers is given in Table VI.

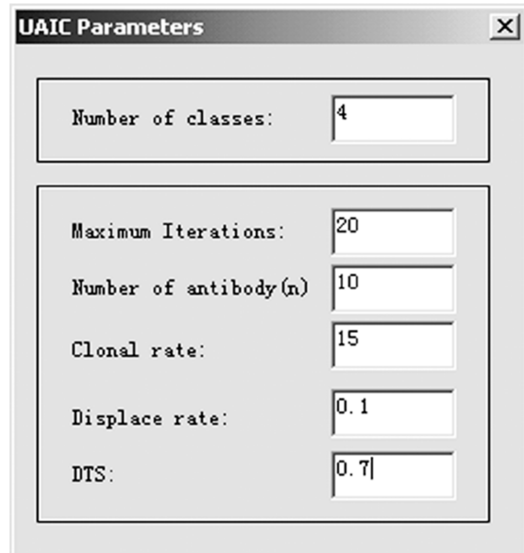


Fig. 11. Parameters dialog in Experiment 2.

As shown in Fig. 12, UAIC and SOM are more capable of distinguishing between the city class and other classes, but SOM is confused by vegetation that is classified to other classes. It is seen from Table VI that the UAIC classifier produces better classification results than traditional classifiers. The details are as follows: UAIC improves overall classification accuracy from 74.71% to 84.63%, an improvement by 9.92% and Kappa coefficient from 0.6631 to 0.7949, an improvement by 0.1318. Based on the above, we can conclude that UAIC is a better classifier for multispectral remote sensing image classification.

C. *Experiment 3: Xiaqiao PHI*

In this experiment, the data are airborne imaging spectrometer (PHI) data, 80 bands taken from Xiaqiao test site which is a mixed agricultural area in China. Eighty bands of PHI image (340×390 pixels) were used in this experiment, and their spectral ranges were from 0.417–0.854 μm. Fig. 13 shows the experimental PHI image. The observed image was expected to fall into seven classes: water, corn1, corn2, road1, road2, soil, and vegetable. The list of classes and the number of labeled samples for each class is given in Table VII.

Fig. 14 shows the values of parameters as set in Experiment 3. It is to be noted that UAIC has a different value of DTS in Experiment 3, from 0.35 in Experiment 1 to 0.85 in Experiment 3. For a convenient comparison between UAIC and traditional unsupervised algorithms, the pixel change threshold is the same value, 3%. Fig. 15(a) illustrates the classification result using UAIC. Fig. 15(b)–(e) illustrates the classification results using K-means, ISODATA, fuzzy K-means, and SOM algorithms. To evaluate the classification accuracy, a test field map is provided

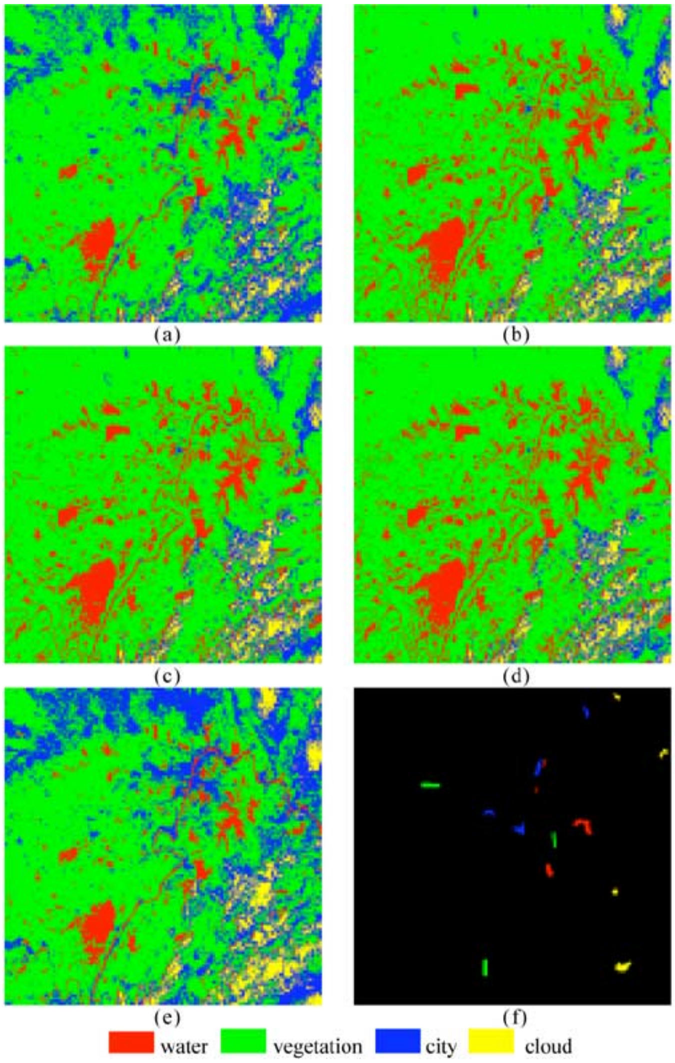


Fig. 12. Unsupervised classification images for Wuhan MODIS image. (a) UAIC. (b) K-means. (c) ISODATA. (d) Fuzzy K-means. (e) SOM (10 × 10 feature map). (f) The image for test fields used in Experiment 2.

TABLE VI
COMPARISON OF FIVE CLASSIFIER PERFORMANCES IN EXPERIMENT 2

Accuracy	K-means	ISODATA	Fuzzy K-means	SOM	UAIC
Overall accuracy	74.71%	74.98%	79.00%	81.95%	84.63%
Kappa coefficient	0.6631	0.6669	0.7199	0.7586	0.7949

in Fig. 15(f) based on the ground truth data. The classification accuracy for the several classifiers is given in Table VIII.

As shown in Table VIII, the UAIC classifier produces better classification results than traditional classifiers. The details are as follows: UAIC improved overall classification accuracy from 70.21% to 81.56%, an improvement by 11.35% and Kappa coefficient from 0.6153 to 0.7535, improving 0.1382. Based on the above, we can make a conclusion that UAIC is the good classifier applied with hyperspectral remote sensing image classification.

Fig. 13. Xiaqiao PHI image RGB(70, 40, 10).

TABLE VII
LIST OF CLASSES AND NUMBER OF LABELED SAMPLES IN EACH CLASS FOR EXPERIMENT 3

Class Name	Number of labeled samples
Water	473
Corn 1	460
Corn 2	478
Road 1	442
Road 2	434
Soil	434
Vegetable	449
Total number of samples	3161

V. SENSITIVITY ANALYSIS OF UAIC

UAIC has two user-defined parameters that significantly influence: 1) the convergence speed; 2) the computational complexity. These parameters are as follows:

1) Clonar_rate: the multiple of clonal antibody
2) DTS: Distance threshold scalar. It affects the number of memory cell population and computational times.

In order to analyze the effects of setting these parameters when running UAIC, Wuhan TM image, shown in Fig. 6, was classified using different values of parameters.

A. Sensitivity in Relation to Parameter Clonal_Rate

In order to study the UAIC sensitivity in relation to Clonal_rate, other parameters were the same with in Experiment 1 and

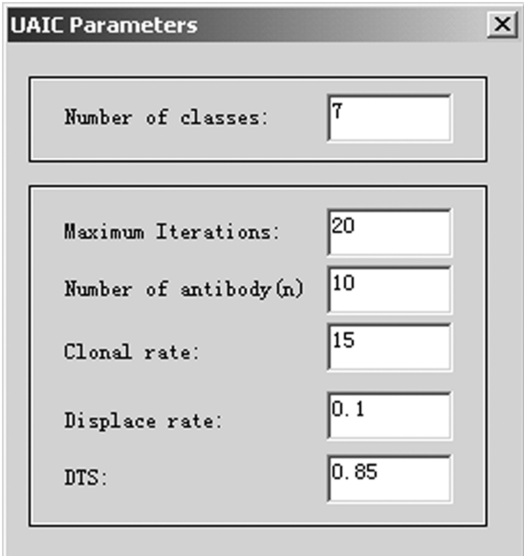


Fig. 14. Parameters dialog in Experiment 3.

Clonal_rate assumed the following values: Clonal_rate $= \{5, 10, 15, 20, 50\}$.

It can be seen from the results presented in Fig. 16 that higher the Clonal_rate, the faster the convergence, in terms of number of generations. However, the computational costs per generation increases linearly with Clonal_rate since the number of antibody population increases linearly.

B. Sensitivity in Relation to Parameter DTS

Distance threshold scalar is very important to maintain the diversity of memory cell population and update the memory cell population. The other parameters are kept the same as in Experiment 1 and DTS assumed the following values: DTS $= \{0, 0.1, 0.2, 0.3, 0.5, 0.7, 0.9, 1\}$. As can be seen from Fig. 17, the number of memory cell decreases from 253 to 4 while DTS increases from 0 to 1.0.

It can be noticed that for value DTS $= 0.0$, the memory cell population is largest and has the better classification results. Nevertheless, it needs much computational time. It is also interesting to observe that the number of memory cell is equal to the number of classes for DTS $= 1.0$.

VI. Conclusion

A novel algorithm based on the paradigm of the nature immune systems, UAIC, was designed and implemented in this paper. The UAIC was successfully applied for classifications of multi/hyperspectral remote sensing images. UAIC was capable of performing data clustering by generating a representative set of memory cells for classification. The key mechanisms and concepts embodied in UAIC include antibody population evolution, clonal selection and memory cell development. To simulate the antibodies in immune systems, we proposed the antibody population model in UAIC and apply the model to the definition of affinity between an antibody and an antigen for remote sensing image classification.

The experimental results consistently show that the proposed UAIC has high classification precision. When compared with

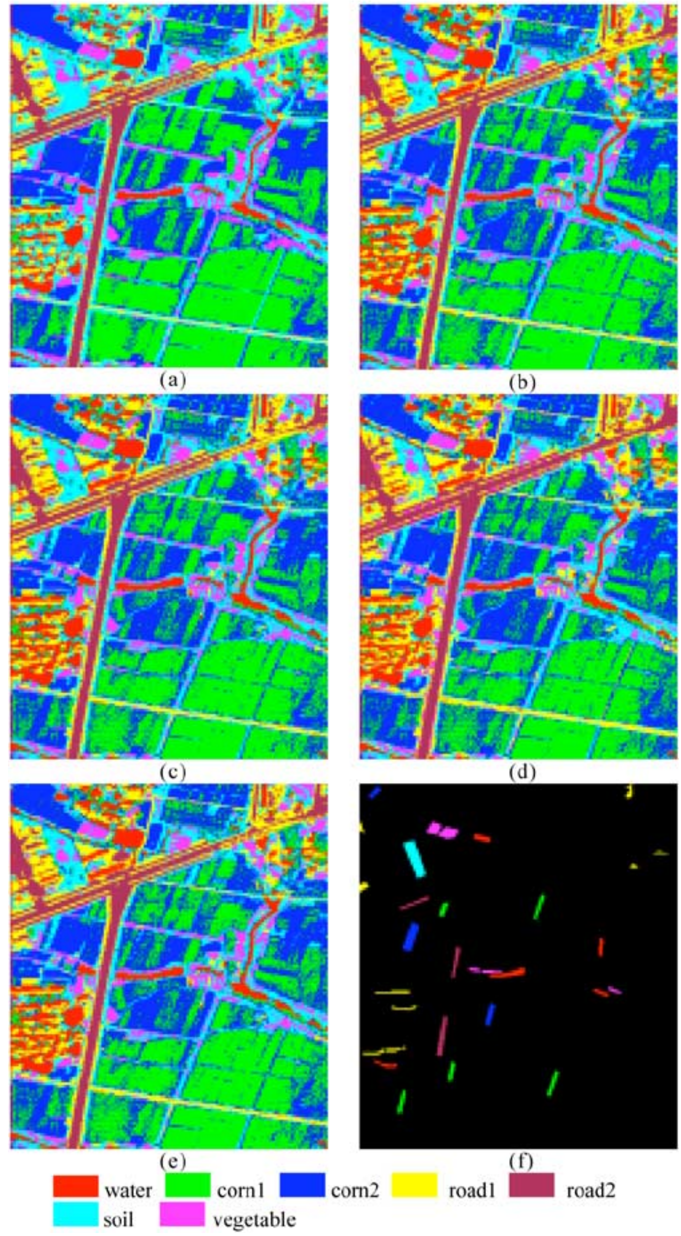


Fig. 15. Unsupervised classification images for Xiaqiao PHI image. (a) UAIC. (b) K-means. (c) ISODATA. (d) Fuzzy K-means. (e) SOM (10 × 10 feature map). (f) The image for test fields used in Experiment 3.

TABLE VIII
Comparison of Five Classifier Performances in Experiment 3

Accuracy	K-means	ISODATA	Fuzzy K-means	SOM	UAIC
Overall accuracy	70.21%	71.33%	75.67%	78.95%	81.56%
Kappa coefficient	0.6153	0.6756	0.7199	0.7177	0.7535

other four unsupervised classifiers, K-means, ISODATA, fuzzy K-means, and SOM, the average performance of UAIC is better than them. In three experiments with different types of images, the average overall accuracy and Kappa coefficient are improved

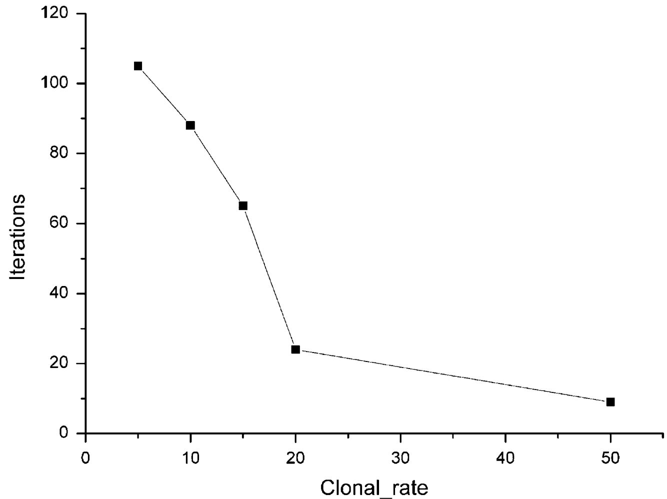


Fig. 16. UAIC sensitivity in relation to Clonal_rate.

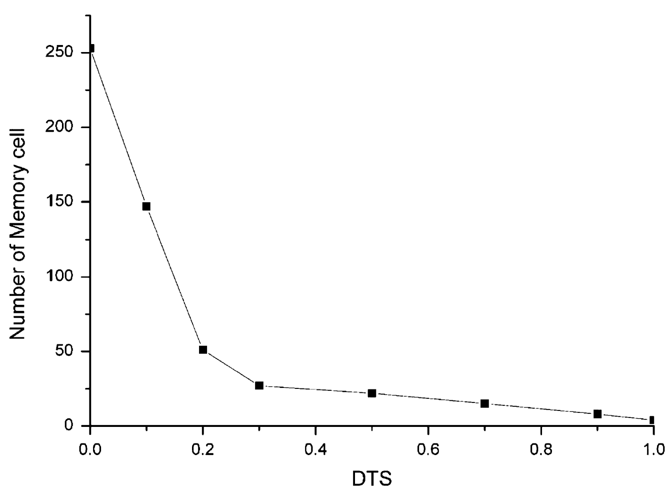


Fig. 17. UAIC sensitivity in relation to DTS.

from 73.08% and 0.655 using K-means algorithm to 83.50% and 0.7794 using UAIC, respectively. This evinces that UAIC is applicable for processing of the multi/hyperspectral remote sensing image and has high classification precision. In future work, we will investigate the sensibility of the proposed method as a function of the number of classes and may use other metrics. Furthermore, we will enhance our classifiers by considering feature selection or extraction using other AIS models in high-dimensional feature space to avoid Hughes phenomena. We may also integrate AIS with a powerful noise remover [33] for improving the classification performance.

References

[1] D. A. Landgrebe, "Hyperspectral image data analysis," *Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[2] R. M. Mohamed and A. A. Farag, "A new unsupervised approach for the classification of multispectral data," in *Proc. 6th Int. Conf. Information Fusion*, vol. 2, Jul. 2003, pp. 951–958.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[4] D. Hall and G. Ball, "Isodata: A novel method of data analysis and pattern classification," Stanford Res. Inst., Stanford, CA, Tech. Rep., 1965.

[5] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Mathematics, Statistics and Probability*, 1967, pp. 281–296.

[6] P. Thitimajshima, "A new modified fuzzy c-means algorithm for multispectral satellite images segmentation," in *Proc. IGARSS*, vol. 4, 2000, pp. 1684–1686.

[7] J. M. Boucher, P. Lena, and J. F. Marchand, "Application of local and global unsupervised Bayesian classification algorithms to the forest," in *Proc. IGARSS*, vol. 2, Aug. 1993, pp. 737–739.

[8] A. H. S. Solberg, T. Taxt, and A. K. Jain, "A Markov random field model for classification of multisource satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 1, pp. 100–113, Jan. 1996.

[9] A. Baraldi and F. Parmiggiani, "A neural network for unsupervised categorization of multivalued input patterns: An application to satellite image clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 305–316, Mar. 1995.

[10] D. Kilpatrick and R. Williams, "Unsupervised classification of Antarctic satellite imagery using Kohonen's self-organizing feature map," in *Proc. IEEE Int. Conf. Neural Networks*, vol. 1, 1995, pp. 32–36.

[11] D. Dasgupta, *Artificial Immune Systems and Their Applications*. Berlin, Germany: Springer-Verlag, 1999.

[12] L. N. de Castro and J. Timmis, *Artificial Immune systems: A New Computational Intelligence Approach*. Berlin, Germany: Springer-Verlag, 2002.

[13] L. N. De Castro and F. J. Von zuben, "Artificial immune systems: Part I—Basic theory and application," State Univ. Campinas, Campinas, Brazil, Tech. Rep. RT DCA 01/99, 1999.

[14] ——, "Artificial immune systems: Part II—A survey of application," State Univ. Campinas, Campinas, Brazil, Tech. Rep. RT DCA 02/0065, 2000.

[15] J. H. Carter, "The immune system as a model for pattern recognition and classification," *J. Amer. Med. Inf. Assoc.*, vol. 7, no. 3, pp. 28–41, 2000.

[16] A. Tarakanov and V. Skormin, "Pattern recognition by immunocomputing," *Proc. 2002 Congress on Evolutionary Computation (CEC)*, vol. 1, pp. 938–943, 2002.

[17] S. A. Hofmeyr and S. Forrest, "Immunity by design: An artificial immune system," in *Proc. Genetic and Evolutionary Computation Conf. (GECCO)*, 1999, pp. 1289–1296.

[18] J. Kim and P. Bentley, "Toward an artificial immune system for network intrusion detection: An investigation of clonal selection with a negative selection operator," in *Proc. Congress on Evolutionary Computation (CEC)*, vol. 2, Seoul, Korea, 2001, pp. 1244–1252.

[19] J. Timmis, M. Neal, and J. E. Hunt, "An artificial immune system for data analysis," *Biosystem*, vol. 55, no. 1/3, pp. 143–150, 2000.

[20] L. N. De Castro and F. J. Von Zuben, "Learning and optimization using the clonal selection principle," *IEEE Trans. Evol. Comput.*, vol. 6, no. 3, pp. 239–251, Jun. 2002.

[21] K. K. Kumar and J. Neidhoefer, "Immunized adaptive critics for level 2 intelligent control," in *Proc. IEEE Int. Conf. Computational Cybernetics and Simulation*, vol. 1, 1997, pp. 856–861.

[22] L. P. Zhang, Y. F. Zhong, and P. X Li, "Applications of artificial immune systems in remote sensing image classification," in *Proc. 20th Congress Int. Soc. for Photogrammetry and Remote Sensing*, vol. 35, Istanbul, Turkey, Jul. 2004, pp. 397–401.

[23] Y. F. Zhong, L. P. Zhang, J. Y. Gong, and P. X. Li, "Classification of multispectral remote sensing image based on clonal selection," *Chinese J. Image Graph.*, vol. 10, no. 1, pp. 18–24, 2005.

[24] N. K. Jerne, "The immune system," *Sci. Amer.*, vol. 229, no. 1, pp. 51–60, 1973.

[25] F. M. Burnet, *The Clonal Selection Theory of Acquired Immunity*. Cambridge, MA, U.K.: Cambridge Univ. Press, 1959.

[26] A. Watkins and L. Boggess, "A new classifier based on resource limited artificial immune systems," *Proc. Congress Evolutionary Computation. Part of the World Congress on Computational Intelligence*, pp. 1546–1551, 2002.

[27] L. N. de Castro, "Comparing immune and neural networks," in *Proc. 7th Brazilian Symp. Neural Networks*, 2002, pp. 250–255.

[28] D. Dasgupta, "Artificial neural networks and artificial immune systems: Similarities and differences," in *Proc. IEEE SMC*, vol. 1, 1997, pp. 873–878.

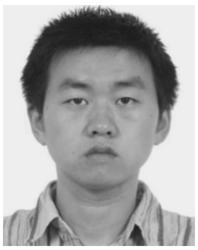
[29] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Trans. Neural Netw.*, vol. 3, no. 5, pp. 698–713, Sep. 1992.

[30] A. Watkins and L. Boggess, "A resource limited artificial immune classifier," in *Proc. 2002 Congress on Evolutionary Computaion, Special Session on Artificial Immune Systems* Honolulu, HI, 2002.

[31] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley, 1990.

[32] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Summaries 4th JPL Airborne Earth Science Workshop* Pasadena, CA, 1992, vol. 92–41, pp. 147–149.

[33] B. Huang, H. G. Li, and X. Huang, "A level set approach to segmentation of oil slicks in SAR images," *Int. J. Remote Sens.*, vol. 26, no. 6, pp. 1145–1156, 2005.



Yanfei Zhong received the B.S. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002. He is currently pursuing the Ph.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan.

His current research interests focus on multi- and hyperspectral image processing, artificial immune systems, and pattern recognition.

Liangpei Zhang received the B.S. degree in physics from Hunan Normal University, ChangSha, China, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1982, 1988, and 1998, respectively.

From 1997 to 2000, he was a Professor in the School of the Land Sciences, Wuhan University. In August 2000, he joined the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, as a Professor and head of the Remote Sensing Section. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence. He is Associate Editor of the *Geo-spatial Information Science Journal*.

Dr. Zhang has served as Co-Chair of the SPIE Series Conferences on Multispectral Image Processing and Pattern Recognition (MIPPR), the Conference on Asia Remote Sensing in 1999, Editor of the MIPPR01, MIPPR05 Symposia, and the Chinese National Committee for the International Geosphere-Biosphere Programme.

Bo Huang is currently an Associate Professor in the Schulich School of Engineering, University of Calgary, Calgary, AB, Canada. He was previously an Assistant Professor in the Department of Civil Engineering, National University of Singapore. His work focuses on the intersection of geospatial information systems, artificial intelligence, operations research, and spatial databases. He has authored and coauthored over 30 papers in refereed international journals and developed several GIS, spatial analysis, and optimization software packages. At present, he serves as a Guest Professor of the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China, and the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing. His current research interests include spatial optimization, spatial statistics for change analysis, and image processing.

Pingxiang Li received the B.S., M.S., and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1986, 1994, and 2003, respectively.

Since 2002, he has been a Professor in the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. His research interests include photogrammetry and SAR image processing.