# Assessment of Performance Improvement in Hyperspectral Image Classification Based on Adaptive Expansion of Training Samples

Hasan Ghasemian*
Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
ghassemi@modares.ac.ir
Maryam Imani
Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
maryam.imani@modares.ac.ir

Received: 07/Jul/2013 Accepted: 12/Apr/2014

**Abstract**
High dimensional images in remote sensing applications allow us to analysis the surface of the earth with more details. A relevant problem for supervised classification of hyperspectral image is the limited availability of labeled training samples, since their collection is generally expensive, difficult and time consuming. In this paper, we propose an adaptive method for improving the classification of hyperspectral images through expansion of training samples size. The represented approach utilizes high-confidence labeled pixels as training samples to re-estimate classifier parameters. Semi-labeled samples are samples whose class labels are determined by GML classifier. Samples whose discriminator function values are large enough are selected in an adaptive process and considered as semi-labeled (pseudo-training) samples added to the training samples to train the classifier sequentially. The results of experiments show that proposed method can solve the limitation of training samples in hyperspectral images and improve the classification performance.

**Keywords:** Classification, Hyperspectral Image, Limited Training Data, Pseudo-Training Samples.

## 1. Introduction

With the development of the remote sensing imaging systems and hyperspectral sensors, the use of hyperspectral image is becoming more interesting. The objective of analysis is to associate each pixel in a hyperspectral image with a proper label. The basis of a classification system is illustrated in Fig. 1 [1]. The increasing of spectral resolution provided by the new sensor technology has brought about new potentials and challenges to data analysis. It is possible to identify more details about classes with higher accuracy than would possible with the data from earlier sensors using a large number of available spectral bands. One the other hand, in order to fully utilize the information contained in the new features, a large number of training samples are needed for using a large number of interesting classes and a large number of available spectral bands. Unfortunately, obtaining of training samples is generally expensive and difficult. When the number of available training samples is relatively small with respect to the number of features, curse of dimensionality problem, Hughes phenomenon occurs [2]. When the given training set is fixed and the dimension of the space grows, the classification accuracy reaches a maximum point and then decreases. When a new feature is added to the data (and the number of training samples is as before) the Bayes error decreased, but at the same time the bias of the classification error increases. This increase is due to the fact that more parameters must be estimated from the same number of available training samples. If the increase in the bias of the classification error is more than the decrease in the Bayes error, then, using of the additional feature degrades the performance of the decision rule. This effect is called the Hughs phenomenon. Therefore, design of the hyperspectral image classifiers that can deal with the small training set has become interesting recently. The suggested solutions can be divided into four categories: 1) fusion of spatial and spectral information [3]-[4];2) feature reduction [5]-[7];3) low complexity classifiers, such as support vector machine (SVM) [8]-[9];4) enlarging the training set by semi-supervised learning. Estimates with smaller covariance matrices can be found by using additional semi-labeled samples. Therefore, better performance can be acquired without the extra cost for selection of more training samples. We focus on the fourth solution in this paper. While the collection of labeled samples is generally time consuming and difficult, unlabeled samples can be generated in a much easier way. Then, the idea of using semi-supervised learning techniques is formed. The main assumption of such methods is that the new training samples can be obtained from a limited set of available labeled samples. A survey of this algorithm is represented in [10]. An adaptive classifier for mitigating the small training sample problem is proposed in [11]. This adaptive classifier that is based on decision fusion, enhances estimation and thus improves the classification accuracy by utilizing the classified samples (referred as semi-labeled samples), in addition to the original training samples. In this classifier, learning of classifier is performed at two steps. At the

* Corresponding Author

beginning of this method, partitioning of observation space is done and several groups of bands are produced. After providing the primary decisions, several rules are used in decision fusion to determine the final label of pixels. In [12], authors used some contextual information such as correlation between a sample and its neighbors for deletion of outlier samples. Then, the semi-labeled pixels are selected from appropriate region. An ensemble algorithm which combines generative model (mixture of Gaussians) and discriminative classifier (support cluster machine) is proposed in [13]. In [14], authors defined a novel composite semi-supervised classifier based on SVMs specifically designed for addressing spectral–spatial categorization of hyperspectral data. In this paper, we use an adaptive classifier for increase the training samples size. Among unlabeled samples whose labels are determined after classification, just samples that their discriminator functions are large enough in Gaussian maximum likelihood (GML) decision rule are selected as semi-labeled samples because they have higher confidence than other classified samples.

The reminder of this paper is organized as follows: some related works about semi-supervised methods are reviewed in section 2. Section 3 describes the suggested approach for increase the training samples size. Section 4 presents the experimental results. Finally, conclusion of this paper is given in section 5.

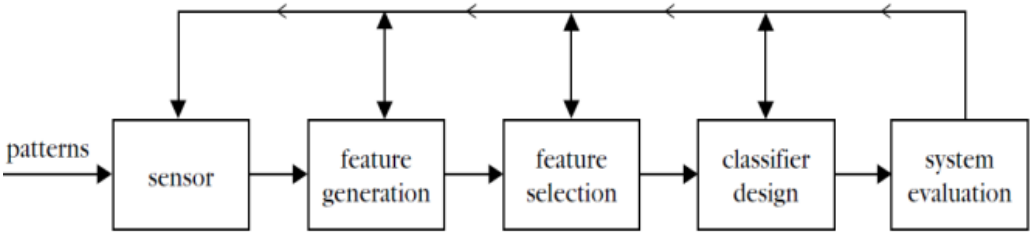


Fig. 1: Basis of classification system [1]

## 2. Some Semi-supervised Approaches

The much information contained in hyperspectral images, allows to characterize and classify the land-covers with more details and improved accuracy and reliability. But the high number of spectral features and the low number of training samples pose the Hughes phenomenon. In the remote sensing applications, many supervised and unsupervised classifiers have been developed to tackle the hyperspectral image classification problem. The main difficulty with supervised method is that performance is basically depends on the quality of training samples. The enough training samples is not available and this is another difficulty. On the other hand, unsupervised methods are not sensitive to the number of training samples, but the relation between clusters and classes is not ensured. Because of represented reasons, it is natural that we use semi-supervised methods for improving of performance. Authors in [15] introduce a semi-supervised graph-based method. Their proposed method has the following characteristics: 1- this method is kernel based and thus the high dimensionality problem is mitigated. 2- The huge number of unlabeled samples is exploited to improve performance. 3- Using graph-based methodology, the relative importance to the labeled samples is considered naturally. 4- The contextual information is incorporated using a family of composite kernels. The graph-based method can be interpreted as a graph $G = (V, E)$ defined on $\mathcal{X} \in \mathbb{R}^N$ ($\mathcal{X}$ denotes a dada set of pixels in a $N$ dimensional space) where the vertex set $V$ is just $\mathcal{X}$ and the edges $E$ are weighted by $W$. The weight matrix $W$ is constructed among all labeled and unlabeled samples and matrix $S$ is defined as follows:

$$S = D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \quad (1)$$

where $D$ is a diagonal matrix with its $(i, i)$ element equal to the sum of the $i$th row of $W$. Given $n$ unlabeled samples, A $n \times c$ matrix $F$ corresponds to a classification on the dada set $\mathcal{X}$ by labeling each point $x_i \in \mathcal{X}$ with a label $y_i = arg\max_{j \le c} F_{ij}$ ($c$ is the number of classes). $F$ can be understood as a vectorial function which assigns a vector $F_i$ to each point $x_i$. A $n \times c$ matrix $Y$ is defined as follows:

$$Y_{ij} = \begin{cases} 1 & y_j = j \\ 0 & y_j \neq j \end{cases} \quad (2)$$

In the proposed graph-based method in [14], following spreading function is iterated until convergence:

$$F(t+1) = \alpha S F(t) + (1-\alpha) Y \quad (3)$$

where $0 < \alpha < 1$ specifies the relative amount of the information from its neighbors and its initial label information.

In semi-supervised methods, if the unlabeled samples are not properly selected, those may confuse the classifier and reduce the classification accuracy. Thus, it is important that the most highly informative unlabeled samples are identified. In [16], a semi-supervised method is proposed that uses self-learning framework. The proposed method in [16] is based on two steps. In the first step, a candidate set, consist of labeled and unlabeled samples, is selected using a self-learning strategy based on spatial information. In the

second step, the standard active learning algorithms on the previously derived candidate set is run to automatically select the most informative samples from the candidate set. Spatial information can be adopted as a reasonable criterion to select unlabeled samples in the proposed method in [16]. First a probabilistic classifier is used to produce a global classification map. Then, the neighbors of the labeled training samples based on a local similarity assumption, are identified and the by analyzing the spectral similarity of spatial neighbors with regard to the original labeled samples, the candidate set is computed. In this method, candidate set is obtained based on spectral and spatial information and its samples are highly reliable. After obtaining candidate set, the most informative unlabeled samples are selected automatically using active learning algorithms and then newly obtained labeled and unlabeled training samples are finally used to retain the classifier. This procedure is repeated iteratively until a convergence criterion.

A semi-labeled bagging technique is proposed by authors in [17]. The novelties of the bagging technique in [17] lie in the definition of a general classification strategy for ill-posed problems by the joint use of training and semi-labeled samples and the design of an effective bagging method (driven by semi-labeled samples) for a proper exploitation of different classifiers based on bootstrapped hybrid training sets. In the bagging algorithms, subsets of bootstrapped samples are generated, and a classifier is built from each subset. The final classification map is obtained by an ensemble rule to achieve a better classification result than the single classifier. The proposed architecture in [17] includes an initial classifier which only exploits the training set to generate the initial classification map. In this way, unlabeled samples become semi-labeled. Then, the generic *b*th classifier of the architecture is defined by selecting semi-labeled samples obtained from the previous classifier. This process is iterated until the desired number of classifiers included in the ensemble of classifiers.

Another semi-supervised method is proposed in [18]. An iterative procedure to produce accurate classification map using a hierarchical segmentation is done. The proposed approach uses the active learning strategies to select the most informative pixels to be labeled. The proposed method in [18] exploits simultaneously the data structures obtained by unsupervised segmentation and information contained in labeled samples. This method uses the available labeled information directly to find the most probable classification map in a hierarchical clustering structure.

## 3. Proposed Adaptive Method for Classification

In this paper, original training samples are samples whose class labels are correctly known and used for training of classifier, i.e. for estimate of mean vectors and covariance matrices in discriminator function of classifier for all classes. Semi-labeled samples that we call them pseudo-training samples are samples whose class labels are determined by a decision rule. They are unlabeled samples before implementation of classification and their class label information partially obtained after classification. The label for a semi-labeled sample can be either right or wrong.

Consider two different estimators $\hat{\theta}, \check{\theta}$ with negligible biases, and assume that $cov(\hat{\theta}) \geq cov(\check{\theta})$. The expected error by using $\hat{\theta}$ is greater than the expected error by using $\check{\theta}$ in the decision rule [1], i.e.

$$E\{\hat{e}\} \geq E\{\check{e}\} \quad (4)$$

By using additional semi-labeled samples, estimates with smaller covariance matrices can be found. If we know which samples have been correctly classified and use them accordingly to re-estimate statistics in addition to original training samples, the estimated statistics should be more precise because the training samples set has been enlarged. In this section, we propose an adaptive method for classification of hyperspectral image using limited training samples. Our used classification method is Gaussian maximum likelihood (GML) that its discriminator function is:

$$g_i(X) = -\ln(|\Sigma_i|) - (X - M_i)^T \Sigma_i^{-1} (X - M_i) \quad (5)$$

where $(M_i)_{n_b \times 1}$ and $(\Sigma_i)_{n_b \times n_b}$ are respectively the mean vector and covariance matrix for class $i$ which are calculated using training samples and $X_{n_b \times 1}$ denotes the unlabeled vector. Also the number of features is denoted by $n_b$. The decision rule of GML is represented by:

$$if\ g_i(X) > g_j(X) \quad for\ all\ j \neq i \quad , i, j \in \{1, 2, \dots, N_c\}$$
$$then\ X \in w_i \quad (6)$$

Where $N_c$ is the number of classes and $X \in w_i$ denotes that $X$ belongs to class $i$. In the proposed method, at first, we classify the pixels of hyperspectral image using limited available training samples. Then, we find pixels that are labeled with high-confidence. These pixels were unlabeled before classification. We call these samples as pseudo- training samples and add them to original training set and use the new extended training set for re-classification in a sequential process.

Our criterion for selection of pseudo-training samples is high-confidence of semi-labeled samples. Assume, two unlabeled samples are labeled after classification and both of them get class label $(x_1, x_2 \in w_i)$. In this conditions, which sample is more appropriate to be considered as pseudo-training sample? The respond is as follows:

$$if\ g_i(x_1) > g_i(x_2), \quad i \in \{1, 2, \dots, N_c\}$$
$$then \quad prob\{x_1 \in w_i\} > prob\{x_2 \in w_i\}$$

where $prob\{\cdot\}$ denotes the probability of $\{\cdot\}$. Therefore, it is reasonable that $x_1$ is selected as a pseudo-training sample. But we want to obtain superlative candidates (high-confidence) for selection of pseudo-training samples. Therefore we define a threshold for discriminator function value. The proposed algorithm for classification of hyperspectral image is illustrated in Fig. 2.

The adaptive proposed method is described as follows:

*Step 1*- The hyperspectral image is classified using just original training samples according Equation (6).

*Step 2*- An appropriate threshold is obtained for selection of pseudo-training samples. We calculate the discriminator function for all training samples in all classes and locate their values in a matrix:

$$G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1N_t} \\ g_{21} & g_{22} & \cdots & g_{2N_t} \\ \vdots & \vdots & \vdots & \vdots \\ g_{N_c1} & g_{N_c2} & \cdots & g_{N_cN_t} \end{bmatrix} \quad (7)$$

Where $N_t$ is the number of training samples per class. An appropriate threshold is selected as follows:

$$Th = \min_{i=1:N_c}\left[\max_{n=1:N_t} g_{in}\right] \quad (8)$$

*Step 3*- selection of pseudo-training samples is done using obtained threshold in step 2 as follows:

$$if \ \ (g_i(X) > g_j(X) \ \ \& \ \ g_i(X) > Th)$$
$$for\ all\ j \neq i \quad , i,j \in \{1,2,\dots,N_c\}$$
$$then\ X\ is\ selected\ as\ high\text{-}confidence\ labeled\ sample\ or\ pseudo\ training\ sample\ for\ class\ i$$

*Step 4*- classification is repeated using the new extended training set (original training samples plus the obtained pseudo-training samples in step 3).

We continue this sequential process to converge to the final values of accuracy and reliability.

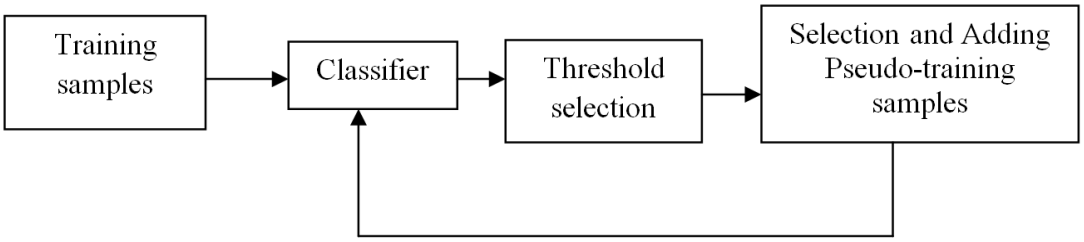


Fig. 2: Proposed algorithm for classification

## 4. Experiments

In order to evaluation of proposed method, several experiments are done. We use three datasets in our experiments. The first data is a synthetic image with size of 80×120. This synthetic scene comprises eight classes and 12 spectral bands which selected from a digital spectral library compiled by the U.S. Geological Survey (USGS) and available online [19]. The false-color image and class map of test data is shown in Fig. 3. The second data is a real multispectral image which is an agricultural segment of Indiana State (F210 dataset) [20]. This image contains 8 different farm classes and is provided in 12 bands with 256 grey levels. The third data is a real hyperspectral image. The hyperspectral data is Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian pines image [20]. This image has 145×145 pixels and contains 16 classes that most of them are different types of crops. The AVIRIS sensor generates 220 spectral bands that we reduced the number of them to 190 by removing 30 absorption and noisy bands. In all experiments 16 random pixels per class are used as original training samples and GML classifier is applied for classification of images. For evaluation of classification performance, we should use some measures. The used measures in this paper are average accuracy, average reliability and kappa

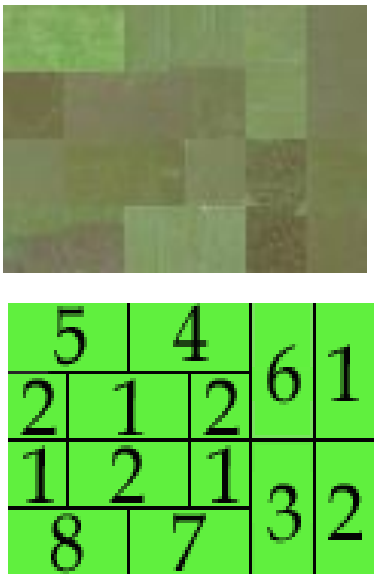


Fig. 3: Test data: (up to down) false -color image, class map.

coefficient (KC). These measures can be represented as follows:

$$average\ ccuracy = mean(P_1, P_2, \dots, P_{N_c}) \quad (9)$$
$$average\ reliability = mean(R_1, R_2, \dots, R_{N_c}) \quad (10)$$

where $mean$ return the mean value of elements. $P_i$ and $R_i$ $i \in \{1,2,\dots,N_c\}$ are defined as follows:

$$P_i = \frac{N_i}{A_{i1}} \ , R_i = \frac{N_i}{A_{i2}} \quad (11)$$

where $N_i$ is the number of test samples of class $i$ that are correctly classified. $A_{i1}$ denotes the total number of test samples that belongs to class $i$ and $A_{i2}$ is the total number of test samples that classified in class $i$. The KC is computed as follows [21]:

$$KC = \frac{N\sum_{c=1}^{n_c} t_{cc} - \sum_{c=1}^{n_c} t_{c+}t_{+c}}{N^2 - \sum_{c=1}^{n_c} t_{c+}t_{+c}} \quad (12)$$

where $N$ denote the number of testing samples and $n_c$ is the number of classes. $t_{cc}$ denotes the number of samples correctly classified in class $c$, $t_{c+}$ is the number of testing samples labeled as class $c$, and $t_{+c}$ is the number of samples predicted as belonging to class $c$.

The obtained values of accuracy and reliability versus the iteration of algorithm in classification of test data are illustrated in Fig. 4. We can see from Fig. 4 that accuracy after 13 iterations and reliability after 12 iterations are converged to their final values. Confusion matrices acquired from test data for GML classifier and proposed classifier are shown in table I and table II respectively. Without any adding pseudo-training samples and just using original training set, 0.76 accuracy and 0.76 reliability were obtained from GML classifier. With using proposed classifier, we can obtain 0.89 accuracy and 0.88 reliability in classification of test data.

Tables III and IV show the confusion matrices for classification of F210 data using GML and proposed classifiers respectively. In this experiment, 0.75 accuracy and 0.59 reliability are acquired from GML classifier. Also 0.83 accuracy and 0.73 reliability are obtained using proposed method. Fig. 5 shows the classification maps using conventional GML classifier and the proposed adaptive classifier for F210 dataset. In both of used dataset (test data and F210 data), the performance of proposed classifier is better than GML classifier considerably.

We repeat the represented experiment for Indian dataset. Because the number of necessary training samples for training the GML classifier is equal to the number of features plus one, we have to reduce the number of features to 15 for using 16 training samples in our experiment. The conventional feature extraction method, maximum noise fraction (MNF), is used for feature reduction. The obtained average accuracy and average reliability versus the number of iterations in proposed adaptive procedure is shown in Fig. 6. As seen from Fig. 6, the accuracy and reliability of classification attain a maximum value after 3 iterations and after that, the performance of classifier is decreased and converges to a less value of efficiency. Doing a feature extraction process before classification of hyperspectral image may cause this behavior for Indian data set that is partially different from multispectral data (test and F210 data sets). Further, if the unlabeled samples are not properly selected, added semi-labeled samples may confuse the classifier. Thus, they may reduce the classification performance. Beside conventional GML classifier, we also compare our proposed method with semi-labeled-sample-driven bagging technique which proposed in [17]. The classification maps using conventional GML classifier, bagging technique and adaptive proposed method are illustrated in Fig. 7.

A summary of classification results which contain average accuracy, average reliability and kappa coefficient for all datasets and three methods (conventional GML classifier, bagging technique [17] and proposed method) are represented in table V. one sees from this table that proposed method has the better performance than conventional GML classifier and the bagging method. The selection of reliable semi-labeled samples with high confidence, which obtained by using a proper threshold, makes our proposed method as an interesting technique for small sample size situation.

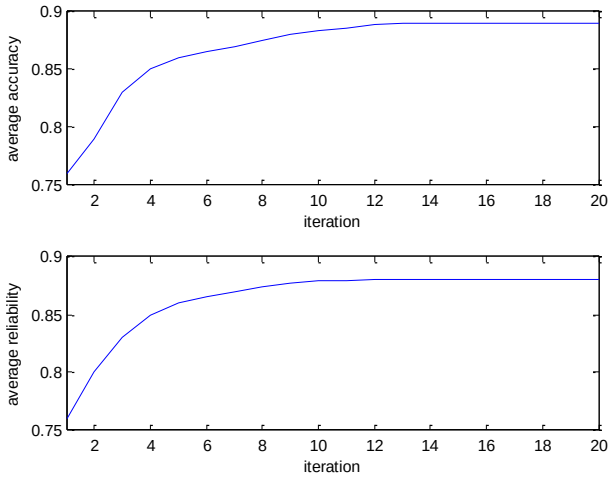


Fig. 4: The average accuracy and average reliability versus the iteration of algorithm (for test data).

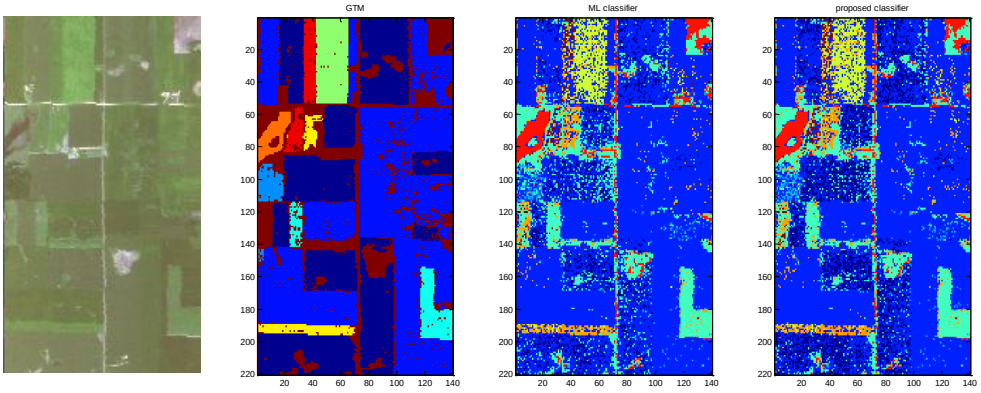


Fig. 5: The result class maps for F210 data set: (Left to right) false -color image, Ground Truth Map (GTM), the result of conventional GML classifier, the result of proposed classifier.

Table I: Confusion Matrix of Conventional GML classifier (test data)

| ML classifier | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 | Class 6 | Class 7 | Class 8 | accuracy |
|---|---|---|---|---|---|---|---|---|---|
| Class 1 | 2133 | 4 | 5 | 69 | 0 | 8 | 1 | 180 | 0.89 |
| Class 2 | 8 | 1940 | 305 | 1 | 7 | 58 | 0 | 81 | 0.81 |
| Class 3 | 2 | 62 | 555 | 33 | 0 | 10 | 1 | 137 | 0.69 |
| Class 4 | 7 | 0 | 1 | 633 | 0 | 75 | 0 | 84 | 0.79 |
| Class 5 | 0 | 0 | 16 | 0 | 736 | 0 | 0 | 48 | 0.92 |
| Class 6 | 0 | 0 | 1 | 15 | 4 | 538 | 70 | 172 | 0.67 |
| Class 7 | 0 | 0 | 0 | 11 | 0 | 407 | 290 | 92 | 0.36 |
| Class 8 | 4 | 0 | 24 | 16 | 2 | 21 | 8 | 725 | 0.91 |
| reliability | 0.99 | 0.97 | 0.61 | 0.81 | 0.98 | 0.48 | 0.78 | 0. 48 | **Average accuracy 0.76** |
| | | | | | | | | | **Average reliability 0.76** |

Table II: Confusion Matrix of proposed classifier (test data)

| Proposed classifier | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 | Class 6 | Class 7 | Class 8 | accuracy |
|---|---|---|---|---|---|---|---|---|---|
| Class 1 | 2291 | 23 | 1 | 45 | 0 | 3 | 0 | 37 | 0.95 |
| Class 2 | 0 | 2342 | 50 | 0 | 0 | 3 | 1 | 4 | 0.98 |
| Class 3 | 7 | 45 | 666 | 48 | 0 | 1 | 0 | 33 | 0.83 |
| Class 4 | 5 | 2 | 1 | 784 | 0 | 8 | 0 | 0 | 0.98 |
| Class 5 | 0 | 0 | 1 | 1 | 787 | 6 | 2 | 3 | 0.98 |
| Class 6 | 0 | 9 | 3 | 32 | 0 | 593 | 159 | 4 | 0.74 |
| Class 7 | 0 | 1 | 0 | 21 | 0 | 215 | 552 | 11 | 0.69 |
| Class 8 | 0 | 3 | 6 | 27 | 0 | 10 | 5 | 749 | 0.94 |
| reliability | 0.99 | 0.97 | 0.91 | 0.82 | 1 | 0.71 | 0.77 | 0. 89 | **Average accuracy 0.89** |
| | | | | | | | | | **Average reliability 0.88** |

Table III: Confusion Matrix of Conventional GML classifier (F210 data)

| ML classifier | **Corn** | **Soybeans** | **Woods** | **Wheat** | **Sudex** | **Oats** | **Pasture** | **Hay** | accuracy |
|---|---|---|---|---|---|---|---|---|---|
| **Corn** | 7110 | 648 | 798 | 248 | 0 | 9 | 0 | 362 | 0.77 |
| **Soybeans** | 61 | 8509 | 1538 | 322 | 358 | 816 | 0 | 233 | 0.72 |
| **Woods** | 27 | 39 | 270 | 12 | 0 | 3 | 1 | 3 | 0.76 |
| **Wheat** | 25 | 25 | 47 | 594 | 0 | 78 | 10 | 45 | 0.72 |
| **Sudex** | 0 | 105 | 3 | 2 | 971 | 102 | 0 | 11 | 0.81 |
| **Oats** | 1 | 70 | 11 | 38 | 35 | 334 | 0 | 54 | 0.62 |
| **Pasture** | 1 | 6 | 6 | 4 | 0 | 4 | 316 | 3 | 0.93 |
| **Hay** | 14 | 52 | 12 | 36 | 15 | 85 | 0 | 439 | 0.67 |
| reliability | 0.98 | 0.90 | 0.10 | 0.47 | 0.70 | 0.23 | 0.97 | 0. 38 | **Average accuracy 0.75** |
| | | | | | | | | | **Average reliability 0.59** |

Table IV: Confusion Matrix of proposed classifier (F210 data)

| Proposed classifier | **Corn** | **Soybeans** | **Woods** | **Wheat** | **Sudex** | **Oats** | **Pasture** | **Hay** | accuracy |
|---|---|---|---|---|---|---|---|---|---|
| **Corn** | 7856 | 917 | 135 | 126 | 0 | 1 | 0 | 139 | 0.86 |
| **Soybeans** | 17 | 10534 | 810 | 14 | 219 | 213 | 0 | 30 | 0.89 |
| **Woods** | 20 | 109 | 210 | 10 | 1 | 1 | 4 | 1 | 0.59 |
| **Wheat** | 2 | 29 | 6 | 714 | 0 | 57 | 15 | 1 | 0.87 |
| **Sudex** | 1 | 12 | 0 | 2 | 1144 | 32 | 0 | 3 | 0.96 |
| **Oats** | 4 | 26 | 0 | 21 | 51 | 417 | 0 | 23 | 0.77 |
| **Pasture** | 0 | 0 | 0 | 1 | 0 | 0 | 338 | 0 | 1 |
| **Hay** | 45 | 37 | 0 | 34 | 38 | 36 | 1 | 462 | 0.71 |
| reliability | 0.99 | 0.90 | 0.18 | 0.77 | 0.79 | 0.55 | 0.94 | 0. 70 | **Average accuracy 0.89** |
| | | | | | | | | | **Average reliability 0.88** |

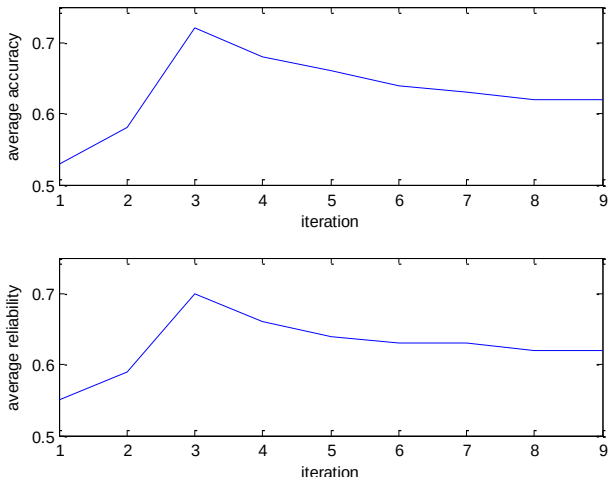


Fig. 6: The average accuracy and average reliability versus the iteration of algorithm for Indian data set

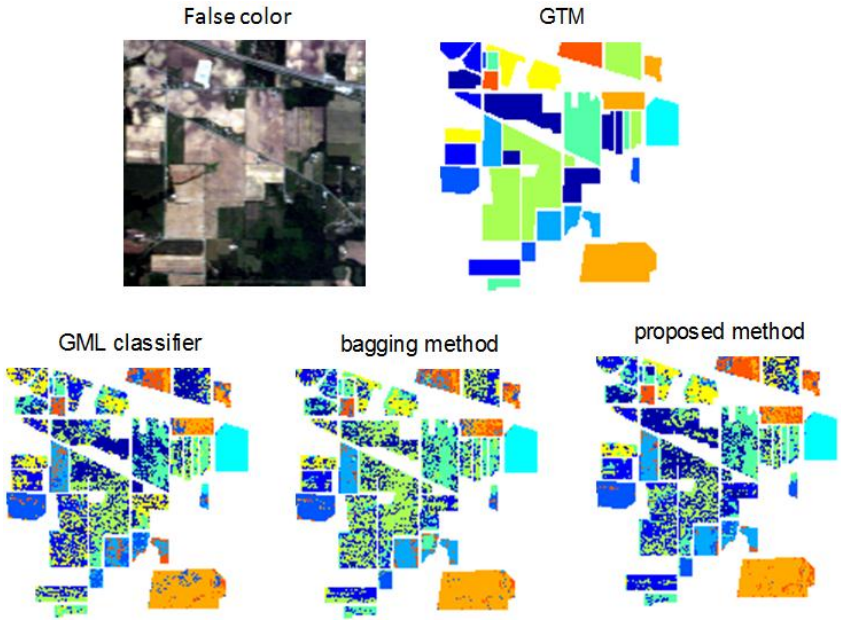


Fig. 7: The result class maps for Indian data set: (Left to right and up to down) false color image, GTM, classification map using GML classifier, classification map using bagging technique, classification map using proposed method.

Table V: Summary of classification results

| dataset | Method | Average accuracy | Average reliability | Kappa coefficient |
|---|---|---|---|---|
| test data | Conventional GML | 0.76 | 0.76 | 0.65 |
| | Bagging technique [17] | 0.82 | 0.81 | 0.73 |
| | Proposed method | 0.89 | 0.88 | 0.78 |
| F210 | Conventional GML | 0.75 | 0.59 | 0.48 |
| | Bagging technique [17] | 0.79 | 0.68 | 0.60 |
| | Proposed method | 0.83 | 0.73 | 0.69 |
| Indian | Conventional GML | 0.62 | 0.60 | 0.54 |
| | Bagging technique [17] | 0.67 | 0.66 | 0.63 |
| | Proposed method | 0.72 | 0.70 | 0.67 |

## 5. Conclusions

For a limited number of available training samples, the classification performance is decreased as the number of features (spectral bands) is increased. This is an important challenge especially in hyperspectral data sets where the ratio of available training samples to dimension of data is small. In this paper, we proposed an adaptive method for classification of hyperspectral images to solve the limitation of available training samples. We select high-confidence labeled samples after primary classification and consider them as semi-labeled

(pseudo-training) samples. The selected pseudo-training samples are added to the original training samples and the new extended training set is used to re-estimate the statistics of classifier. This process is continued sequentially until the accuracy and reliability of classifier converge to the final values for multispectral data or gain the maximum accuracy and reliability (the best possible performance) for hyperspectral images. Our experiment results show that proposed method has better performance than conventional GML classifier and semi-labeled-sample-driven bagging technique. The proposed method can be an effective solution to cope with the small sample size situation.

## References


[1] Q. Z. Jackson and D. Landgrebe, "Design of an Adaptive Classification Procedure for The Analysis of High-dimensional Data With Limited Training Samples," TR-ECE 01-5, Purdue University, West Lafayette, Indiana, Dec. 2001.

[2] G. F. Hughes, "On the mean accuracy of statistical pattern recognition," IEEE Trans. Inf. Theory, Vol. IT-14, No. 1, Jan. 1968, pp. 55–63.

[3] T. Achalakul and S. Taylor, "Real-time multispectral image fusion," Concurr. Computat.: Pract. Exper., Vol. 13, No. 12, Sep. 2001, pp. 1063–1081.

[4] A. Keshavarz, H. Ghassemian1, and H. Dehghani, "Hierarchical classification of hyperspectral images by using SVMs and "same class neighborhood property"," IEEE Symposium on Geoscience and Remote Sensing, Vol. 5, July 2005, pp. 3219 – 3222.

[5] A.C. Jensen and A.S. Solberg, "Fast hyperspectral feature reduction using piecewise constant function approximations," IEEE Trans. Geosci. Remote Sens. Lett., Vol 4, No. 4, Oct. 2007, pp. 547–551.

[6] L. O. Jimenez and D. A. Landgrebe, "Hyperspectral data analysis and feature reduction via projection pursuit," IEEE Trans. Geosci. Remote Sensing, Vol. 37, Nov. 1999, pp. 2653–2667.

[7] C. Cariou, K. Chehdi, and S. Le Moan, "BandClust: An Unsupervised Band Reduction Method for Hyperspectral Remote Sensing," IEEE Trans. Geosci. Remote Sens. Lett., Vol. 8, No. 3, MAY 2011, pp. 565-569.

[8] C.-H. Li, B.-C. Kuo, C.-T. Lin, and C.-S. Huang, "A Spatial–Contextual Support Vector Machine for Remotely Sensed Image Classification," IEEE Trans. Geosci. Remote Sensing, Vol. 50, No. 3, March 2012.

[9] F. Bovolo, L. Bruzzone, and Lorenzo Carlin, "A Novel Technique for Subpixel Image Classification Based on Support Vector Machine," IEEE Trans. Image Proces. Vol. 19, No. 11, Nov. 2010, pp. 2983 - 2999.

[10] X. Zhu, "Semi-supervised learning literature survey," Comput. Sci., Univ. Wisconsin-Madison, Madison, WI, Tech. Rep. 1530, 2005.

[11] H. Dehghani and H. Ghassemian, "Adaptive Classification of Hyperspectral Image," 12th Iranian Conference on Electrical Engineering, 11-13 May, 2004.

[12] A. Keshavarz and H. Ghassemian, "An Improving in Estimate of Classifier Parameters for Hyperspectral Images Using Semi-labeled Samples and Deletion of Outlier Samples", 16th Iranian Conference on Electrical Engineering, 13-15 May, 2008, pp. 306-311, (Printed in Persian).

[13] M. Chi, Q. Kun, J.-A. Benediktsson, and Rui Feng, "Ensemble Classification Algorithm for Hyperspectral Remote Sensing Data," IEEE Trans. Geosci. Remote Sens. Lett, Vol. 6, No. 4, oct. 2009, pp. 762-766.

[14] M., G. Camps-Valls, and L. Bruzzone, "A Composite Semisupervised SVM for Classification of Hyperspectral Images," IEEE Trans. Geosci. Remote Sens. Lett, Vol. 6, No. 2, April 2009, pp. 234-238.

[15] G. Camps-Valls, T. Bandos Marsheva, and D. Zhou, "Semi-Supervised Graph-Based Hyperspectral Image Classification," IEEE Trans. Geosci. Remote Sensing, Vol. 45, No. 10, Oct. 2007, pp. 3044 - 3054.

[16] I. Dopido, J. Li, P.R. Marpu, A. Plaza, J.M. Bioucas Dias, and J.A. Benediktsson, "Semisupervised Self-Learning for Hyperspectral Image Classification," IEEE Trans. Geosci. Remote Sensing, Vol. 51, No. 7, July 2013, pp. 4032-4044.

[17] M. Chi and L. Bruzzone,"A semilabeled-sample-driven bagging technique for ill-posed classification problems," IEEE Geosci. Remote Sensing lett., Vol. 2, No. 1, Jan. 2005, pp. 69-73.

[18] J. Munoz- Mari, D. Tuia, and G. Camps- Valls, "Semisupervised Classification of Remote Sensing Images With Active Queries," IEEE Trans. Geosci. Remote Sensing, Vol. 50, No. 10, Oct. 2012, pp. 3751 - 3763.

[19] http://speclab.cr.usgs.gov/spectral-lib.html.

[20] D. A. Landgrebe, Signal Theory Methods in Multispectral Remote Sensing, Hoboken, NJ: Wiley, 2003.

[21] J. Cohen,"A coefficient of agreement from nominal scales," Edu. Psychol. Meas., Vol. 20, No. 1, 1960, pp. 37–46.



**Hasan Ghasemian** received the B.S.E.E. degree from Tehran College of Telecommunication in 1980 and the M.S.E.E. and Ph.D. degree from Purdue University, West Lafayette, USA in 1984 and 1988 respectively. Since 1988, he has been with Tarbiat Modares University in Tehran, Iran, where he is a Professor of Computer and Electrical Engineering. His research interests focus on Multi-Source Signal/Image Processing and Information Analysis and Remote Sensing.

**Maryam Imani** received the B.Sc. and M.Sc. Degrees in Electrical Engineering from Shahed University, Tehran, Iran in 2009 and 2011 respectively. She is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering at the Tarbiat Modares University, Tehran, Iran. Her research interests include the hyperspectral image analysis, feature extraction and pattern recognition in remote sensing applications.