CrossMark

A novel approach for change detection of remotely sensed images using semi-supervised multiple classifier system

Moumita Roy [a], Susmita Ghosh [a], Ashish Ghosh [b],*

[a] *Department of Computer Science and Engineering, Jadavpur University, Kolkata, India*
[b] *Center for Soft Computing Research, Indian Statistical Institute, Kolkata, India*

A R T I C L E   I N F O

A B S T R A C T

In this article, a novel approach using ensemble of semi-supervised classifiers is proposed for change detection in remotely sensed images. Unlike the other traditional methodologies for detection of changes in land-cover, the present work uses a multiple classifier system in semi-supervised (leaning) framework instead of using a single weak classifier. Iterative learning of base classifiers is continued using the selected unlabeled patterns along with a few labeled patterns. Ensemble agreement is utilized for choosing the unlabeled patterns for the next training step. Finally, each of the unlabeled patterns is assigned to a specific class by fusing the outcome of base classifiers using some combination rule. For the present investigation, multilayer perceptron (MLP), elliptical basis function neural network (EBFNN) and fuzzy *k*-nearest neighbor (*k*-nn) techniques are used as base classifiers. Experiments are carried out on multi-temporal and multi-spectral images and the results are compared with the change detection techniques using MLP, EBFNN, fuzzy *k*-nn, unsupervised modified self-organizing feature map and semi-supervised MLP. Results show that the proposed work has an edge over the other state-of-the-art techniques for change detection.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Change detection is a process of detecting temporal effects of multi-temporal images [1,2]. This process is used for finding out changes in land covers over time by analyzing remotely sensed images of a geographical area captured at different time instants. Changes can occur due to natural hazards (e.g., disaster, earthquake), urban growth, deforestation, etc. [1–5]. Change detection is one of the most challenging tasks in the field of pattern recognition and machine learning [6].

Change detection can be viewed as an image segmentation problem, where two groups of pixels are to be formed, one for the changed class and the other for the unchanged one. Process of change detection can be broadly classified into two categories: supervised [7–9] and unsupervised [10–17]. Supervised techniques have certain advantages like they can explicitly recognize the kind of changes occurred and are robust to different atmospheric and light conditions of acquisition dates. Various methodologies exist in the literature to carry out supervised change detection, e.g., post classification [1,8,18], direct multi-date classification [1], kernel based methods [9], etc. Having several advantages, applicability of supervised methods in change detection is poor due to mandatory requirement of sufficient amount of ground truth information, collection of which is expensive, hard and monotonous too. On the contrary, in unsupervised approach [10–17], there is no need of

---

* Corresponding author. Tel.: +91 33 2575 3110/3100; fax: +91 33 2578 3357.
  E-mail address: ash@isical.ac.in (A. Ghosh).

additional information like ground truth. Due to depletion of labeled patterns, unsupervised techniques seem to be compulsory for change detection. Unsupervised change detection process can be of two types: context insensitive (spectral based) [1,12] and context sensitive (spatial based) [10,11,13–16,19].

In change detection, it may so happen that the category information of a few labeled patterns could be collected easily by experts [20]. However, if the number of these labeled patterns is small, then this information may not be sufficient for developing any supervised method. In such a scenario, knowledge of labeled patterns, though not much in amount, may be completely unutilized if unsupervised approach is carried out. Under this circumstance, semi-supervised approach [21,22] can be opted instead of unsupervised or supervised ones. Semi-supervision uses a small amount of labeled patterns with abundant unlabeled ones for learning, and integrates the merits of both supervised and unsupervised strategies to make full utilization of the collected patterns. Semi-supervision has been used successfully for improving the performance of clustering and classification [23–26] when sufficient amount of labeled data are not present.

Semi-supervised approaches were explored for the use of multiple classifier system (MCS) [27–30]. Many applications in real life domains, i.e., change detection, medical image analysis, face recognition suffer from the problem of unavailability of labeled information. Therefore, semi-supervised MCS are required and have been studied in past [27–30] (see Table 1). As to the knowledge of the authors, no such applications exists in change detection domain using semi-supervised MCS. This motivated us to explore the capacity of ensemble classifier embedded with semi-supervision framework to improve the performance of change detection process when a few labeled patterns are available.

In the proposed method, merits of both semi-supervised learning and ensemble learning are integrated in a single platform for detecting changes from remotely sensed images. The traditional algorithms [1,8,9,18,31] for change detection is mainly relaying on a single classifier in either supervised or semi-supervised framework. Unlike this, in the present work, a set of semi-supervised classifiers is used for change detection. In the present investigation, multilayer perceptron (MLP) [32], elliptical basis function neural network (EBFNN) [32–34] and fuzzy k-nearest neighbor techniques (k-nn) [35] are used as the base classifiers.

To assess the effectiveness of the proposed method, experiments are carried out on two multi-temporal and multi-spectral images of Mexico area and Island of Sardinia. The present study concludes that the proposed semi-supervised MCS is better suited for the task of change detection than the other state-of-the-art techniques.

The rest of the article is organized into four sections. Section 2 describes the proposed methodology. Description of the data sets used to carry out the investigation is provided in Section 3. In Section 4, implementation details and experimental results are discussed. Conclusion is drawn in Section 5.

2. The proposed algorithm

In the present work, an ensemble of semi-supervised classifiers is proposed for change detection. The contribution of the present work is twofold: at first an algorithm is designed to integrate semi-supervised learning and ensemble learning in a single platform and then the proposed algorithm is used for the betterment of change detection process when a few labeled patterns are available. Unlike the other state-of-the-art techniques in the literature of semi-supervised multiple classifier system (i.e. co-training [29], tri-training [30], co-forest [44]), the proposed algorithm during the iterative learning process utilizes the agreement between all the networks in the ensemble for collecting the most confident labeled patterns.

Table 1
Tabular representation for state-of-art techniques.

Application area	Technique and tools
Water quality prediction	Greedy ensemble selection family of algorithms for ensembles of regression models was used for searching the best subset of regressors by taking some local greedy decisions [36]
Image recognition	A hybrid approach was used by combining type-2 fuzzy logic, modular neural networks and the Sugeno integral [37]
Sentiment classification	A detailed comparative study for sentiment classification using ensemble techniques was carried out. Here, ensemble of classifiers were designed in three different level of combination, i.e., classification level, feature level and combination level [38]
Prediction of chaotic time series	Ensemble of adaptive network based fuzzy inference system was used. Average and weighted average combination rules were applied for final decision [39]
Cardiac arrhythmia Classification	fuzzy K-Nearest Neighbors, Multi Layer Perceptron with Gradient Descent and momentum Backpropagation, and Multi Layer Perceptron with Scaled Conjugate Gradient Backpropagation were used as base classifiers. Finally, a Mamdani type fuzzy inference system was used as a combiner [40]
Hyperspectral image classification	Combination of two discriminative classifiers: sparse multinomial logistic regression and quadratic discriminant analysis were used. Initially, both the classifiers were trained using a few labeled samples. Then, the set of unlabeled samples was collected for next training step by combining the estimation obtained by both classifiers [41]
Classification of multi-annual remote sensing data	Ensemble of classifiers (i.e., random forests) was trained on a multi-spectral image captured from an agricultural region. It was adapted to classify the image from another year in semi-supervised framework [42]
Remotely sensed image segmentation	Co-training strategy was used under variational Bayesian framework. There were two disjoint feature sets and each of them was used by a Gaussian mixture model. Co-training strategy in a bootstrap mode was utilized for estimation of parameters [43]

As already mentioned, multilayer perceptron (MLP) [32], elliptical basis function neural network (EBFNN) [32–34] and fuzzy k-nearest neighbor techniques (k-nn) [35] are used as the base classifiers. Here, a few labeled patterns are required for semi-supervised learning. These labeled patterns can be collected in many ways. In the proposed method, for experimental purpose, an equal amount of labeled patterns, from both the classes (changed and unchanged), are picked up randomly from the ground truth. For the labeled patterns, the target values, support values and the membership values are assigned to either $[1, 0]$ or $[0, 1]$ depending on their class labels. Detailed description of the proposed change detection technique is presented in subsequent sections.

2.1. Generation of input pattern

The difference image $DI = \{l_{mn}, 1 \leqslant m \leqslant p, 1 \leqslant n \leqslant q\}$ is produced by the Change vector analysis technique [1] from two co-registered and radiometrically corrected γ-spectral band images Y_1 and Y_2, each of size $p \times q$, of the same geographical area captured at different times T_1 and T_2. Here, gray value of the difference image DI at spatial position (m, n), denoted as l_{mn}, is calculated as,

$$l_{mn} = (int)\sqrt{\sum_{i=1}^{\gamma} \left(l_{mn}^i(Y_1) - l_{mn}^i(Y_2) \right)^2},$$

(1)

where $l_{mn}^i(Y_1)$ and $l_{mn}^i(Y_2)$ are the gray values of the pixels at the spatial position (m, n) in the ith band of the images Y_1 and Y_2, respectively.

From the difference image DI, the input pattern for a particular pixel position is generated by considering the gray value of the said pixel as well as those of its neighboring ones to exploit (spatial) contextual information from neighbors. In the present methodology, 2nd order neighborhood system [31] is used. Here, each input pattern consists of two features: (i) gray value of its own and (ii) average of the gray values of its neighboring pixels including its own value.

The y-dimensional input pattern of the (m, n)th pixel position of DI is denoted by $\overrightarrow{X_{mn}} = [x_{mn,1}, x_{mn,2}, \ldots, x_{mn,y}]$. Here, a mapping algorithm is used to normalize the feature values of the input pattern in [0, 1]. The ith feature value ($i = 1, 2, \ldots, y$) of the y-dimensional input pattern, $\overrightarrow{X_{mn}}$, is normalized as

$$x_{mn,i} = \frac{x_{mn,i} - c_{min}}{c_{max} - c_{min}},$$

(2)

where, c_{max} and c_{min}, respectively, are the maximum and the minimum gray values of the DI.

2.2. Support value estimation using EBFNN

Elliptical basis function neural network [34] uses the full covariance matrix to improve the performance of conventional radial basis function network with a few basis functions. The network consists of three layers: one input layer, one hidden layer and one output layer. The center of each basis function is initialized by the mean value of the labeled patterns from the corresponding class and the center are kept fixed for each training step. There are as many neurons in the hidden layer as the number of basis functions and no weighted connection is present between the neurons in the input layer and the hidden layer. In the present work, only two basis functions are used corresponding to both the changed and unchanged classes. The network is trained by updating the weights using the Least Mean Square algorithm (LMS) [32] to minimize the error between the target value (or soft class label) and the predicted output value of input patterns. Initially, the network is trained by a few labeled patterns only.

After convergence of the training of the network, each of the unlabeled patterns is passed through the EBFNN (trained) to predict the output values for both the classes. Let, $\mu r(m, n) = [\mu r_1(m, n), \mu r_2(m, n)]$ be the two degrees of support, estimated by the EBFNN classifier, where $\mu r_1(m, n)$ and $\mu r_2(m, n)$ are the support values of the (m, n)th pattern in the unchanged and changed classes, respectively. The output value of the (m, n)th unlabeled pattern for the ith class (here, $i = 1$ or 2), $yr_i(m, n)$ is assigned to $\mu r_i(m, n)$. The unlabeled pattern is more likely to belong to the class for which its support value is higher.

2.3. Support value estimation using MLP

MLP [31,32] has one input layer, one output layer and one or more hidden layers. MLP is trained by updating the weight using backpropagation algorithm [32] to minimize the error between the target value and the predicted output value of the patterns. Let a be the number of layers in the network and the jth neuron in the rth layer ($r \geqslant 1$) receives total input of the form $vm_j^r(m, n)$ from $(r - 1)$th layer. Then, the jth neuron in the rth layer ($r \geqslant 1$) produces an output of the for $ym_j^r(m, n) = \frac{1}{(1 + exp(-vm_j^r(m, n)))}$. Initially, the connection weight is updated by using the labeled patterns only.

After convergence, each of the unlabeled patterns is tested by the trained MLP to predict the output values for both the changed and the unchanged classes. Let, $\mu m(m, n) = [\mu m_1(m, n), \mu m_2(m, n)]$ be the two degrees of support, estimated by the network, where $\mu m_1(m, n)$ and $\mu m_2(m, n)$ are the support values of the (m, n)th pattern in the unchanged and changed classes, respectively. The output value of the (m, n)th unlabeled pattern for the ith class (here, $i = 1$ or 2), $ym_i^{(a-1)}(m, n)$ is assigned to $\mu m_i(m, n)$.

2.4. Support value estimation using fuzzy k-nn

Fuzzy k-nn classifier assigns membership values in both the classes for each of the unlabeled patterns. Here, these membership values are treated as the predicted support values. Let, $\mu k(m,n) = [\mu k_1(m,n), \mu k_2(m,n)]$ be the two degrees of support of the (m,n)th unlabeled pattern, where $\mu k_1(m,n)$ and $\mu k_2(m,n)$ are the membership values, which are estimated by the classifier for the unchanged and changed classes, respectively. For each unlabeled pattern, its k nearest labeled patterns are determined. To search for the k number of nearest neighbors, instead of using all the labeled patterns, we considered only those which lie within a neighborhood (window) around that unlabeled pattern. This reduces time requirement for searching. Let W be the set of k nearest labeled patterns of the (m,n)th unlabeled pattern. The membership value $\mu k_i(m,n)$ of the (m,n)th unlabeled pattern in the ith class (here, $i = 1$ or 2) is calculated as

$$\mu k_i(m,n) = \frac{\sum_{\overrightarrow{X_{ef}} \in W} \mu k_i(e,f)(1/\|\overrightarrow{X_{mn}} - \overrightarrow{X_{ef}}\|^{2/(fm-1)})}{\sum_{\overrightarrow{X_{ef}} \in W} (1/\|\overrightarrow{X_{mn}} - \overrightarrow{X_{ef}}\|^{2/(fm-1)})}; \tag{3}$$

where fm is a parameter, called fuzzifier, which determines the weighting factor of the distance to control the neighbor's contribution to the membership value.

2.5. Estimation of soft class label using 'maximum' combination rule

First, the support values (or, output values) for each of the unlabeled patterns are obtained with the base classifiers using only a few labeled patterns (same labeled patterns for all the base classifiers). Here, the two degrees of support for a particular pattern (m,n) using the three base classifiers are organized into a matrix, called decision profile ($DP(m,n)$) [45]. The matrix $DP(m,n)$ is represented as follows:

$$DP(m,n) = \begin{pmatrix} d_{11}(m,n) & d_{12}(m,n) \\ d_{21}(m,n) & d_{22}(m,n) \\ d_{31}(m,n) & d_{32}(m,n) \end{pmatrix} = \begin{pmatrix} \mu r_1(m,n) & \mu r_2(m,n) \\ \mu m_1(m,n) & \mu m_2(m,n) \\ \mu k_1(m,n) & \mu k_2(m,n) \end{pmatrix}. \tag{4}$$

Now, the soft class labels (or, target value) for each of the unlabeled patterns are calculated using 'maximum' combination rule [46] on these support values. The soft class label is assigned to the unlabeled patterns because we do not want to commit about the class label at this moment. The target value of the (m,n)th unlabeled pattern in the ith class ($target_i(m,n)$) is calculated as follows:

$$target_i(m,n) = \max\{d_{1i}(m,n), d_{2i}(m,n), d_{3i}(m,n)\}. \tag{5}$$

Here, the estimated target value in the ith class for the (m,n)th unlabeled pattern is normalized as $target_i(m,n) = target_i(m,n)/\sum_{i=1}^{2} target_i(m,n)$ to make the summed up target values in the two classes to 1.

2.6. Selection of unlabeled patterns for the next training step

After this, most confident unlabeled patterns are selected for the next training step. Here a method has been suggested to obtain a set of most confident unlabeled patterns (denoted as, U). The (m,n)th unlabeled pattern is selected as the most confident pattern for the ith class (here, $i = 1$ or 2) and placed into the set U, if it satisfies all the following conditions:

  (i) the computed support value in the ith class is always larger than those which are obtained for other classes for the pattern in case of all the base classifiers;
  (ii) the support value in the ith class is always greater than α_i for the said pattern for all the base classifiers; and
  (iii) the absolute difference between the estimated target values of the pattern in both the classes is greater than β_i.

By conditions (i) and (ii), high agreement between all the base classifiers has been taken into consideration. Condition (iii) is used to avoid selection of confusing unlabeled patterns.

To increase diversity of the base classifiers, three mutually exclusive sets (of the same size) for the unlabeled patterns (denoted as, UR, UM and UK) for (iterative) training of each of the base classifiers are generated by randomly selecting the most confident unlabeled patterns from the set of U.

2.7. Semi-automatic computation of the parameters α_i and β_i

After estimating soft class label values for unlabeled patterns, the most confident unlabeled patterns for the ith class are collected using two important parameters, i.e. α_i and β_i. In the present work, a technique has been suggested for (semi)-automatic estimation of these parameters. For computing both the parameters, at the onset, a set of unlabeled patterns are selected for which the estimated soft class label value is maximum in ith class.

Then, for calculating the value of α_i, average, minimum and maximum (denoted by $avgt_i$, $mint_i$ and $maxt_i$, respectively) of the estimated soft class label in ith class for the selected unlabeled patterns are computed. Now, if the average value is nearer to the minimum value then it implies that the estimated soft class label value for most of the unlabeled patterns in the ith class are closer to the minimum value. In this case, if the $alpha_i$ is assigned higher value than the average, then there is a high chance of selection of a few number of most-confident unlabeled patterns. In this situation, the value of $alpha_i$ is fixed at an average value. On the other hand, to avoid too much selection, if the average value is nearer to maximum value, then $alpha_i$ is kept fixed at $(avgt_i + maxt_i)/2$ to set a higher value than the average.

In case of obtaining the β_i, the difference between the estimation soft class label value in both the classes is computed for each of the selected unlabeled patterns in ith class. Now β_i is fixed at average of the computed difference value for ith class.

2.8. Iterative learning of ensemble classifiers until convergence

After the selection of most confident unlabeled patterns, learning of EBFNN and MLP are then carried out again using the labeled patterns along with the unlabeled patterns (from the UR and UM, respectively). For the next training step, the centers, co-variance matrices and smoothness parameters of all the basis functions for EBFNN are also updated using the unlabeled patterns (in the set UR) along with a few labeled patterns.

Now, the support values of all the unlabeled patterns in both the classes are re-estimated using EBENN, MLP and fuzzy k-nn classifier.

In case of fuzzy k-nn classifier, the membership values in both the classes for all the unlabeled patterns are estimated again considering the labeled patterns as well as the unlabeled patterns from the set UK. Here, the membership values for the selected unlabeled patterns in both the classes are assigned using the estimated target values of the unlabeled patterns in the previous training step. In the present investigation, it has been noticed that the membership values of the selected unlabeled patterns participate to estimate the membership values of the other unlabeled patterns, as well as, at the same time its own membership values are also estimated using the other selected unlabeled patterns along with the labeled patterns. Therefore, the membership values for the selected unlabeled patterns are updated after calculating the membership values for all the unlabeled patterns.

At the end of each training step, itr, the sum of square error (ξ_{itr}) between the estimated target values at itrth and $(itr - 1)$th training steps is calculated as,

$$\xi_{itr} = \sum_{m=1,n=1}^{p,q} \sum_{i=1}^{2} \left(tfinal_i^{itr}(m,n) - tfinal_i^{(itr-1)}(m,n) \right)^2; \tag{6}$$

where the estimated target value of the (m,n)th pattern in ith class at the itrth training step, $target_i(m,n)$ is stored in $tfinal_i^{itr}(m,n)$.

Iterative learning of the ensemble classifier, re-estimation of the target values of the unlabeled patterns, updating the value of α_i and β_i for the ith class, selection of the unlabeled patterns are continued until the difference between ξ_{itr} and $\xi_{(itr+1)}$ of two consecutive training steps becomes less than ϵ (where ϵ is a small positive quantity) or the number of training steps exceeds a prespecified value. After convergence, the hard class labels are assigned to the unlabeled patterns depending on their target values. The algorithmic representation of the proposed methodology is given in Table 2.

3. Description of data sets

To establish the effectiveness of the proposed methodology, experiments are carried out on two multi-temporal remotely sensed images corresponding to the geographical areas of Mexico and Sardinia Island of Italy.

Table 2
Algorithmic representation of the proposed work.

Step 1:	Pick up a few labeled patterns from the reference map
Step 2:	Estimate the support values (or, output values) in both the classes for each of the unlabeled patterns by the three base classifiers using a few labeled patterns only
Step 3:	Estimate soft class label (or, target value) for the unlabeled pattern by applying 'maximum' combination rule on the support values in both the classes, obtained by the three base classifiers
Step 4:	Compute the parameters α_i and β_i used for collecting the most confident unlabeled patterns in ith class
Step 5:	For the next training step, select the most confident unlabeled patterns using ensemble agreement of all the base classifiers
Step 6:	Estimate the support values in both the classes for the unlabeled patterns by the three base classifiers using the labeled patterns as well as the selected unlabeled patterns
Step 7:	Repeat Steps 3, 4, 5 and 6 until convergence. At convergence, goto Step 8
Step 8:	Assign a hard class label to each of the unlabeled patterns
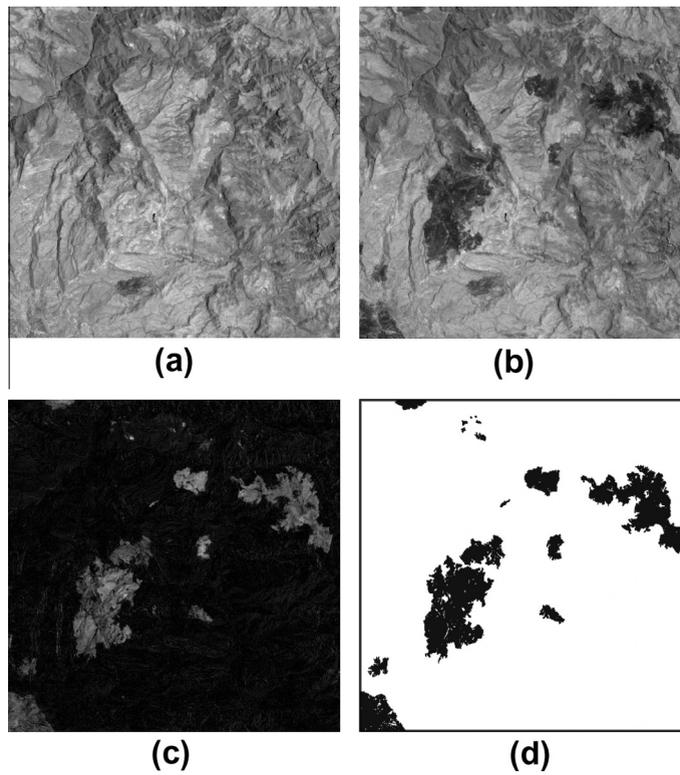
Fig. 1. Images of Mexico area. (a) Band 4 image acquired in April 2000, (b) band 4 image acquired in May 2002, (c) corresponding difference image generated by CVA technique, and (d) a reference map of the changed area.
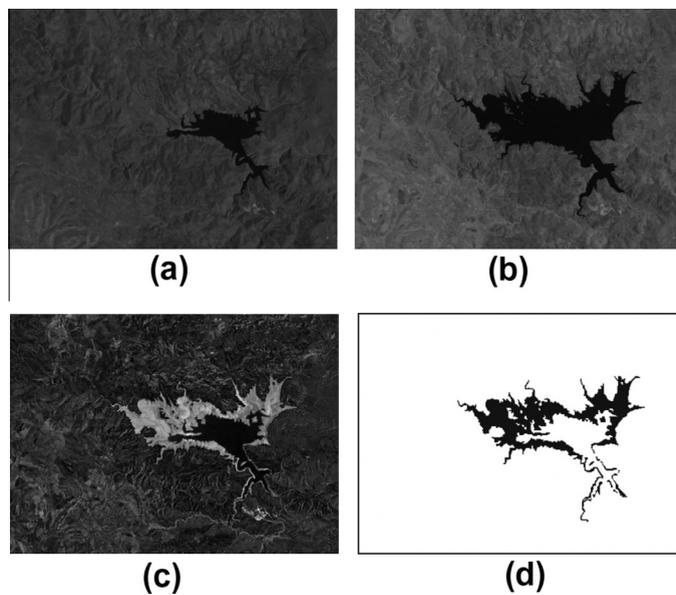
Fig. 2. Images of Sardinia Island, Italy. (a) Band 4 image acquired in September 1995, (b) band 4 image acquired in July 1996, (c) difference image generated by CVA technique using bands 1, 2, 4, and 5, and (d) a reference map of the changed area.

3.1. Data set related to Mexico area

This data set consists of two multi-spectral images of the Landsat-7 satellite captured by the Landsat Enhanced Thematic Mapper Plus (ETM+) sensor over an area of Mexico acquired on 18th April, 2000 and 20th May, 2002. From the entire

Table 3
Results on Mexico data set.

Techniques used	Training patterns (%)	Avg. *MA*	Avg. *FA*	Avg. *OE*	Avg. micro F_1	Avg. macro F_1	Avg. *Kappa*
Supervised	0.1	1532.8823	2393.9608	3926.8431	0.9581	0.9585	0.9163
EBFNN	0.2	1388.2549	2328.1961	3716.4510	0.9604	0.9607	0.9209
	0.3	1325.2941	2085.9216	3411.2157	0.9636	0.9638	0.9272
Supervised	0.1	1696.3725	1328.0588	3024.4314	0.9670	0.9671	0.9340
MLP	0.2	1415.2941	1420.1372	2835.4314	0.9693	0.9694	0.9386
	0.3	1333.1176	1430.0588	2763.1765	0.9701	0.9702	0.9403
Supervised	0.1	2874.6667	831.5294	3706.1961	0.9583	0.9590	0.9166
fuzzy	0.2	2224.6470	960.5882	3185.2353	0.9647	0.9649	0.9294
k-nn	0.3	2292.9216	895.3921	3188.3137	0.9646	0.9649	0.9300
Semi-	0.1	1561.8627	1136.5294	2698.3922	0.9705	0.9706	0.9411
supervised	0.2	1385.8823	1178.2157	2564.0980	0.9721	0.9722	0.9443
MLP	0.3	1296.8823	1200.8627	2497.7451	0.9729	0.9729	0.9458
Unsupervised MSOFM	–	1366.7619	1623.6190	2990.3810	0.9679	0.9678	0.9356
Proposed semi-	0.1	1430.5294	1186.4706	2617.0000	0.97153	0.9716	0.9431
supervised	0.2	1212.8431	1230.8039	2443.6471	0.9735	0.9736	0.9471
MCS	0.3	1183.8431	1177.9412	2361.7843	0.9744	0.9744	0.9489

Table 4
Results on Sardinia data set.

Techniques used	Training patterns (%)	Avg. *MA*	Avg. *FA*	Avg. *OE*	Avg. micro F_1	Avg. macro F_1	Avg. *Kappa*
Supervised	0.1	1659.0588	488.4314	2147.4902	0.9157	0.9193	0.8315
EBFNN	0.2	1201.5294	581.4314	1782.9608	0.9338	0.9348	0.8676
	0.3	1088.4706	628.0784	1716.5490	0.9370	0.9377	0.8741
Supervised	0.1	1165.8431	576.0000	1741.8431	0.9354	0.9365	0.8709
MLP	0.2	1072.2353	649.1765	1721.4118	0.9369	0.9377	0.8739
	0.3	1051.1961	651.6863	1702.8823	0.9378	0.9384	0.8756
Supervised	0.1	2256.8431	248.3333	2505.1765	0.8937	0.9025	0.7881
fuzzy	0.2	1572.6470	326.8431	1899.4902	0.9259	0.9286	0.8519
k-nn	0.3	1153.2353	358.7843	1512.0196	0.9431	0.9441	0.8861
Semi-	0.1	1290.7843	422.7451	1713.5294	0.9351	0.9367	0.8703
supervised	0.2	1268.3137	399.2549	1667.5686	0.9370	0.9384	0.8741
MLP	0.3	1255.5098	391.3333	1646.8431	0.9379	0.9391	0.8758
Unsupervised MSOFM	–	1070.3810	578.4762	1648.8571	0.9398	0.9395	0.8790
Proposed semi-	0.1	925.5490	217.2157	1142.7647	0.9573	0.9582	0.9146
supervised	0.2	864.6471	242.1961	1106.8431	0.9589	0.9596	0.9178
MCS	0.3	837.5098	246.1569	1083.6667	0.9599	0.9605	0.9198

available Landsat scene, a section of 512×512 pixels are selected as test site. A fire destroyed a large portion of the vege-
tation in the considered region between two acquisition dates. Fig. 1(a) and (b), respectively show the band 4 images of April,
2000 and May, 2002. The difference image (Fig. 1(c)) created by spectral band 4 using CVA technique is only used for further
analysis. To evaluate the performance of the algorithms, a reference map (Fig. 1(d)) is used. The reference map contains
25,599 changed and 236,545 unchanged pixels.

3.2. Data set related to Sardinia Island, Italy

Two multi-spectral images are acquired in September, 1995 and July, 1996 by the Landsat Thematic Mapper (TM) sensor
of the Landsat-5 satellite. The test site of 412×300 pixels of a scene includes the lake Mulargia on the Sardinia Island, Italy.
The water level of the lake increased between two acquisition dates (lower central part of the image reflects the same).
Fig. 2(a) and (b), respectively, show the 1995 and 1996 images of band 4. CVA technique has been applied on spectral bands
1, 2, 4, and 5 of these two images to obtain the difference image (Fig. 2(c)). 7480 changed and 116,120 unchanged pixels are
present in the reference map (Fig. 2(d)).

4. Results and analysis

As mentioned in Section 1, to evaluate the effectiveness of the proposed method, experiments are conducted on two mul-
ti-temporal remotely sensed images. Performance of the proposed technique is compared with those of some supervised
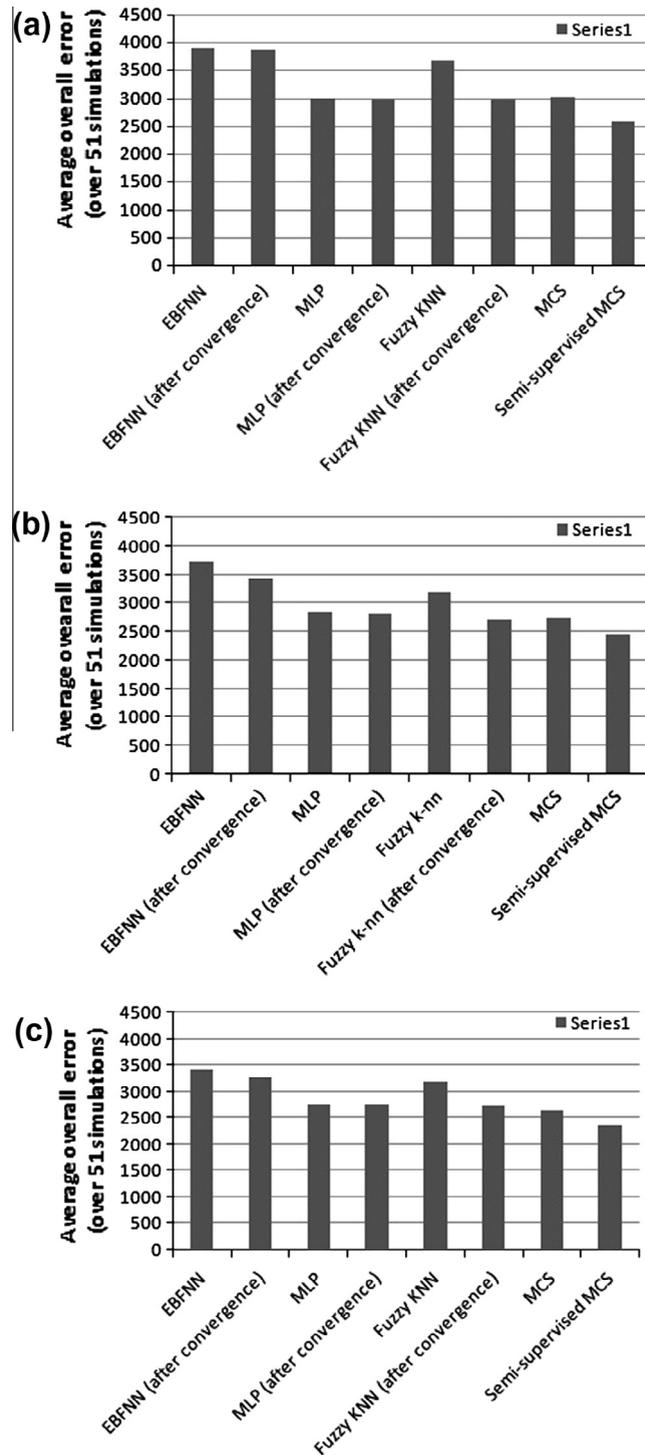
Fig. 3. Bar charts for Mexico data set. (a) with 0.1% training patterns, (b) with 0.2% training patterns, and (c) with 0.3% training patterns.

methods based on MLP [32], EBFNN [32–34] and fuzzy *k*-nn [35], an unsupervised method based on MSOFM [11] and a semi-supervised method based on MLP [31]. For experimentation, three different percentages of training patterns (0.1%, 0.2%, and 0.3%) are considered and 51 simulations are conducted in each case. Due to computing resource limitation we have conducted 51 simulations.

Here, ensemble of three models is used instead of considering two of them. There is a clear intuitive reason behind this, which also corroborates our experimental findings during the present investigation. If we consider two models, it may so
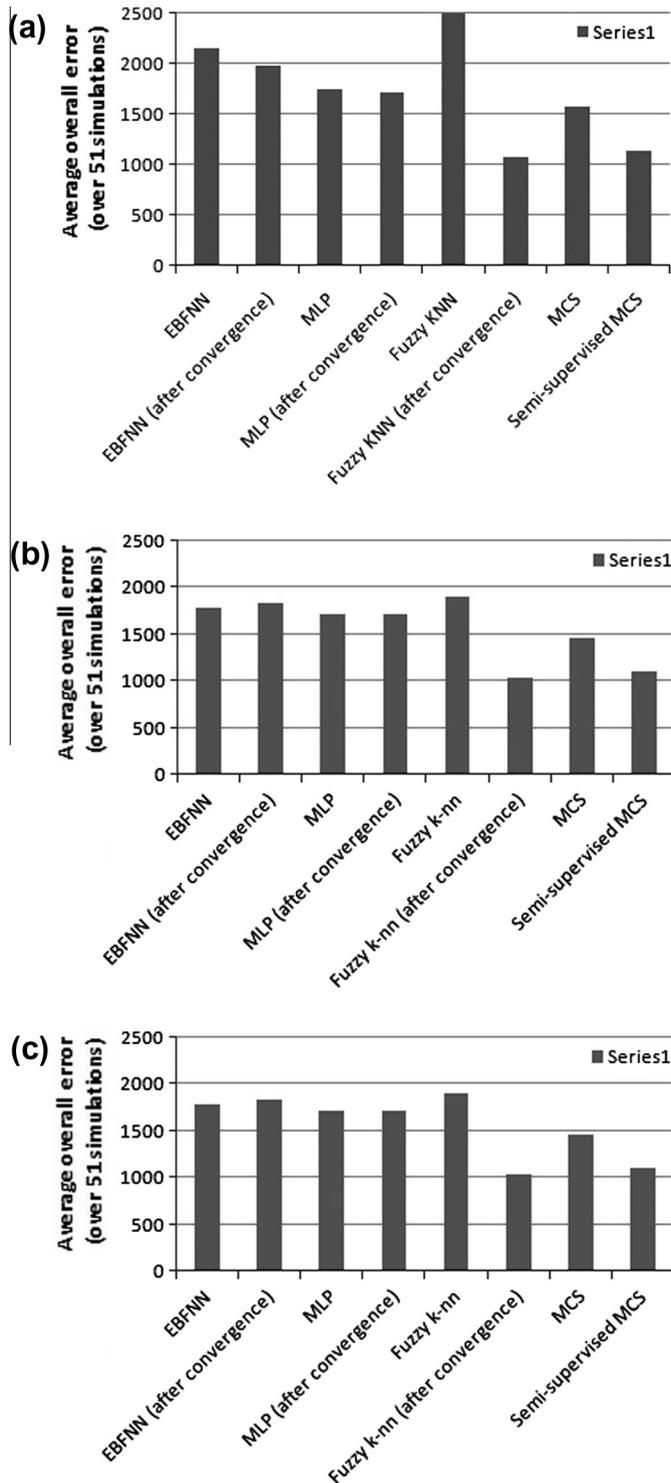
Fig. 4. Bar charts for Sardinia data set. (a) with 0.1% training patterns, (b) with 0.2% training patterns, and (c) with 0.3% training patterns.

happen that they make contradictory decisions about a pixel and it becomes very hard to make any final decision using the combination rules. For this reason, most of the ensemble architectures consider odd number of models. In ensemble archi-tecture, if the number of (diverse) classifiers is increased, then the final decision becomes more robust and errorfree, but it will also increase the complexity of the methodology. Under the implementation point of view, it may be better to fix the number within a limit. Under this circumstance, in the present investigation, we considered three models, MLP, EBFNN

and fuzzy k-nn. Here, the base classifiers, which provide graded response for each pattern in both the classes (changed and unchanged) are preferred. It is also noticed that if we consider any two of them then the performance (in terms of all the measuring indices) of the proposed methodology is noticeably degraded.

To find out the k nearest neighbors for each unlabeled pattern, it has been experimentally observed that the required window size should decrease with increase in the percentage of labeled patterns. For Mexico data set, window size was taken as $251 \times 251, 201 \times 201$ and 151×151 using 0.1%, 0.2% and 0.3% labeled patterns, respectively. For Sardinia data set, the same was taken $91 \times 91, 71 \times 71$ and 61×61 using 0.1%, 0.2% and 0.3% labeled patterns, respectively. The value of k was fixed to 9 and 5 for Mexico and Sardinia data set, respectively. The *fm* was fixed to 2 for all the data sets. In the present work, since local window based fuzzy k-nn approach is considered, a situation might occur when sufficient (here, k) number of labeled patterns are not present within the window around an unlabeled pattern. In this scenario, for the unlabeled pattern, the membership values in both the classes are assigned to 0. In the present investigation, experiments are carried out with different combiners, i.e. 'average', 'product', 'minimum' and 'maximum'. Among the four combiners, the 'maximum' combination rule is selected as the best one for the proposed methodology and it is used for typical illustration.

To assess the effectiveness of the proposed methodology, various performance measuring indices are considered in our investigation and these are as follows: number of missed alarms (*MA*), number of false alarms (*FA*), number of overall error (*OE*), micro-averaged F_1 measure (MicroF_1) [47], macro averaged F_1 measure (MacroF_1) [47] and Kappa measure (*Kappa*) [48]. The average values (over 51 simulations) of all the performance measuring indices are considered for comparative analysis. Results for Mexico and Sardinia data sets are put in Tables 3 and 4, respectively. The bar charts of average overall error
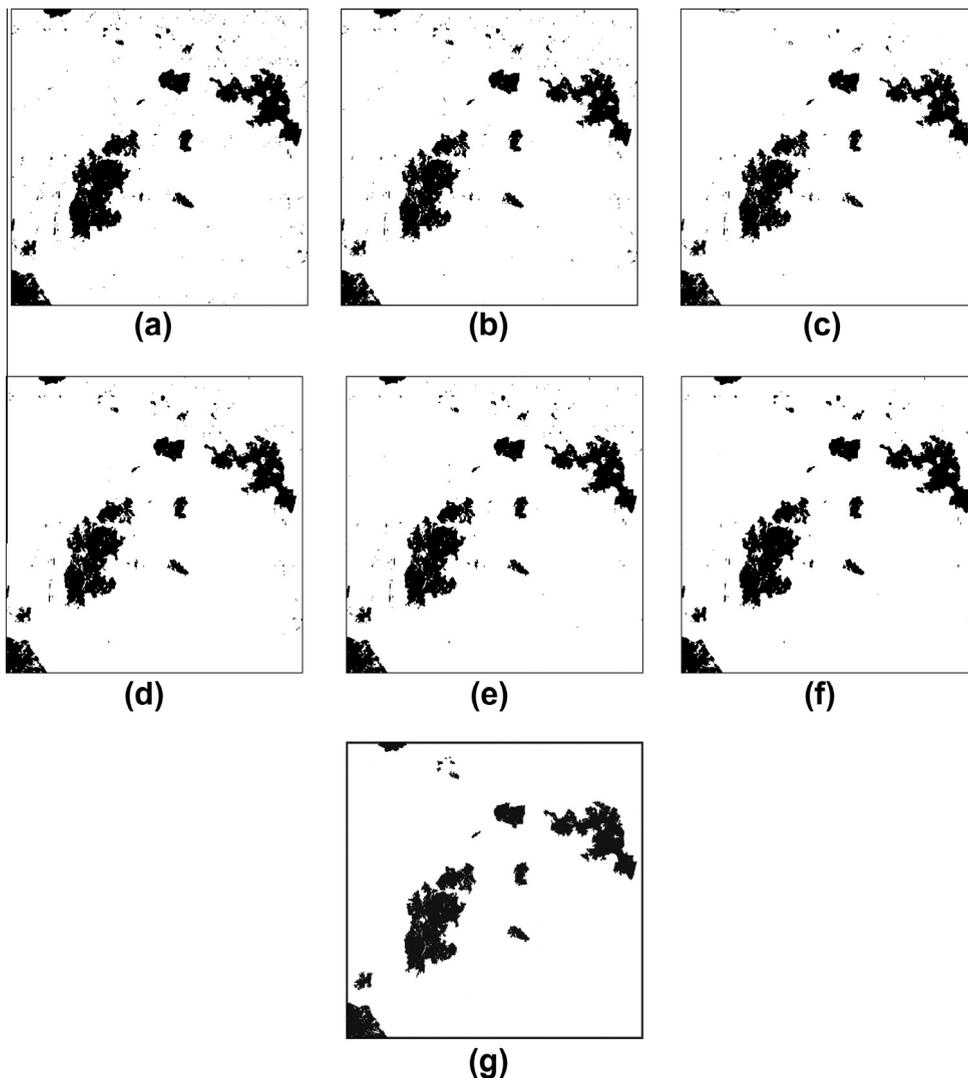
Fig. 5. Change detection maps obtained for Mexico data set by: (a) supervised EBFNN, (b) supervised MLP, (c) supervised fuzzy k-nn, (d) semi-supervised MLP, (e) the proposed semi-supervised MCS (with 0.1% training pattern), (f) unsupervised MSOFM, and (g) the reference map.

for Mexico and Sardinia data sets with different percentage of training patterns are also depicted in Figs. 3 and 4, respectively. In the charts, abbreviated names of the techniques are depicted along *x*-axis and the corresponding values of average overall error are depicted along *y*-axis. Here the average overall errors are obtained using EBFNN, MLP, fuzzy *k*-nn classifier are put in the bar charts. Ensemble of these three base classifiers (MCS) applying maximum combination rule with a few labeled patterns are displayed in the charts. The average overall error for each of the base classifiers and the proposed semi-supervised MCS, after the convergence of iterative (semi-supervised) learning are also plotted in the bar charts.

From Table 3, it has been noticed for Mexico data set that the proposed method (with three different percentage of training patterns used) outperforms the corresponding methods using EBFNN [32–34] and MLP [32] for all the six measuring indices. It has also been clearly observed from the table that the number of missed alarms which are obtained using EBFNN and MLP with 0.2% and 0.3% training patterns, are similar with those obtained using the semi-supervised MCS with 0.1% and 0.2% training patterns. It has also been observed that the results obtained using the proposed technique are significantly better than those found out using fuzzy *k*-nn classifier [35] and the semi-supervised method based on MLP [31] in almost all the cases except the case of false alarms. In comparison to semi-supervised MLP, the betterment is more prominent with increase in percentage of training patterns. Among all six performance measuring indices with three different percentages of training patterns used, it has been noticed that the results obtained using the proposed technique are better than those obtained using the unsupervised method based MSOFM [11] in most of the cases except the case of missed alarms with 0.1% training patterns.

By analyzing the results depicted in Table 4, it has been seen for Sardinia data set that the proposed methodology (in terms of all the measuring indices) outperforms all the five approaches used for comparison in the present investigation. It has been observed that the number of overall error (i.e. 1083) obtained using 0.3% training patterns (the amount is very less) is a great achievement from the viewpoint of change detection.

From Figs. 3 and 4, it has been observed for Mexico and Sardinia data sets that the proposed method has an edge over the conventional ensemble classifier. For the Mexico data set, it has been noticed that the performance of all the individual base classifiers are improved with the semi-supervised leaning; and the performance of semi-supervised ensemble classifier is significantly better than those of the individual base classifiers after convergence. But this scenario is not true for all the data sets. It has been observed in Fig. 4 that the performance of the proposed semi-supervised ensemble classifier for Sardinia
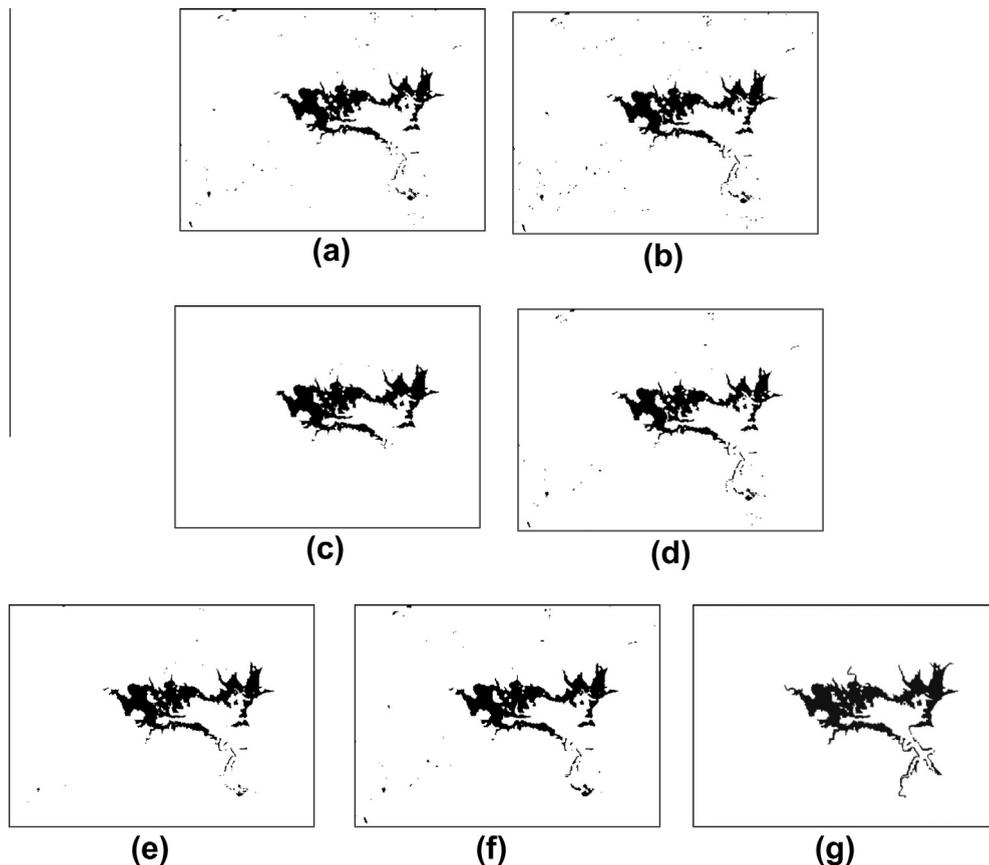
Fig. 6. Change detection maps obtained for Sardinia data set by: (a) supervised EBFNN, (b) supervised MLP, (c) supervised fuzzy *k*-nn, (d) semi-supervised MLP, (e) the proposed semi-supervised MCS (with 0.1% training pattern), (f) unsupervised MSOFM, and (g) the reference map.

data set is significantly better than those of the supervised methods and the ensemble classifier with a few labeled patterns. After the convergence of iterative (semi-supervised) learning, it has been noticed that the performance of one (i.e., fuzzy k-nn) of the base classifiers is improved more than that of the corresponding MCS. Therefore, the improvement in the performance of the individual base classifiers after convergence of semi-supervised learning using the selected unlabeled patterns cannot be predicted beforehand.

For visual illustration, the change detection maps are displayed in figures, generated by considering majority voting on the change detection maps, obtained over 51 different simulations, using EBFNN, MLP, fuzzy k-nn, semi-supervised MLP, unsupervised MSOFM, and the proposed semi-supervised MCS. Here, majority voting principle assigns a pattern to a particular class if it was assigned to the said class in maximum number of simulations. The change detection maps obtained using these six approaches for Mexico and Sardinia data sets are shown, correspondingly, in Figs. 5 and 6. From Fig. 5, it has been observed that the changed region present in upper left corner of the reference map (in Fig. 5(g)) is wrongly identified as the unchanged one (i.e. missed alarms) in the map obtained using fuzzy k-nn technique (in Fig. 5(c)); whereas the region is correctly identified in the map obtained using the proposed method (in Fig. 5(e)). It has been also noticed that small changed regions are scattered all over the maps (in Fig. 5(a) and (b)) obtained using EBFNN and MLP, which are basically wrongly identified ones (see the reference map in Fig. 5(g)). These scattered and wrongly identified changed regions (i.e. false alarms) are comparatively less in the change detection maps obtained using the proposed method (in Fig. 5(e)). It has been also noticed for Sardinia data set that the scattered and wrongly identified changed regions are significantly less (i.e. false alarms) in the map (in Fig. 6(e)) obtained using the proposed method than those obtained using the other five approaches (in Fig. 6(a)–(d) and (f)). From Fig. 6, it has also been found that the edge of the changed region present in lower right corner in the reference map (in Fig. 6(g)) is totally absent (i.e. missed alarms) in the maps obtained using the fuzzy k-nn technique (in Fig. 6(c)); whereas this changed region is partially identified in the map obtained using the proposed approach (in Fig. 6(e)). From these observations, it has been concluded that the change detection maps obtained using the proposed semi-supervised MCS more accurately resemble the reference map.

5. Conclusion

In the present work, a novel technique for change detection is proposed using semi-supervised multiple classifier system (MCS). Here, instead of using a single weak learner (classifier under inadequacy of labeled patterns), a set of classifiers is used and they are trained in semi-supervised learning framework by exploiting the unlabeled patterns along with a few labeled patterns. For collecting the unlabeled patterns, ensemble agreement between all the participating base classifiers is utilized. In the present investigation, multilayer perceptron, elliptical basis function neural network and fuzzy k-nearest neighbor techniques are used as the base classifiers.

Experiments are carried out on multi-temporal and multi-spectral data sets to confirm the effectiveness of the proposed technique. From the results, it has been found that the proposed semi-supervised ensemble approach is better suited for change detection than the other state-of-the-art techniques when a small amount of labeled patterns is available. In future we will perform a detailed study about the dependency of the performance of the proposed approach on the selection of the parameters and the type of scene used for application.

Acknowledgments

References

[1] A. Singh, Digital change detection techniques using remotely-sensed data, Int. J. Remote Sens. 10 (6) (1989) 989–1003.
[2] M.J. Canty, Image Analysis, Classification and Change Detection in Remote Sensing: with Algorithms for ENVI/IDL, CRC Press, Taylor & Francis Group, Boca Raton, 2007.
[3] Q. Zhang, J. Wang, X. Peng, P. Gong, P. Shi, Urban built-up land change detection with road density and spectral information from multi-temporal Landsat TM data, Int. J. Remote Sens. 23 (15) (2002) 3057–3078.
[4] K.R. Merril, L. Jiajun, A comparison of four algorithms for change detection in an urban environment, Remote Sens. Environ. 63 (2) (1998) 95–100.
[5] L. Bruzzone, D.F. Prieto, An adaptive parcel-based technique for unsupervised change detection, Int. J. Remote Sens. 21 (4) (2000) 817–822.
[6] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, USA, 2006.
[7] F. Yuan, K.E. Sawaya, B.C. Loeffelholz, M.E. Bauer, Land cover classification and change analysis of Twin cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing, Remote Sens. Environ. 98 (2005) 317–328.
[8] G.M. Foody, Monitoring the magnitude of land-cover change around the southern limits of the Sahara, Photogramm. Eng. Remote Sens. 67 (2001) 841–847.
[9] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Mari, J.L. Rojo-Álvarez, M. Martinez-Ramón, Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection, IEEE Trans. Geosci. Remote Sens. 46 (6) (2008) 1822–-1835.
[10] S. Ghosh, L. Bruzzone, S. Patra, F. Bovolo, A. Ghosh, A context-sensitive technique for unsupervised change detection based on Hopfield-type neural networks, IEEE Trans. Geosci. Remote Sens. 45 (3) (2007) 778–789.

[11] S. Ghosh, S. Patra, A. Ghosh, An unsupervised context-sensitive change detection technique based on modified self-organizing feature map neural network, Int. J. Approx. Reason. 50 (1) (2009) 37–50.

[12] F. Melgani, G. Moser, S.B. Serpico, Unsupervised change detection methods for remote sensing images, Opt. Eng. 41 (12) (2002) 3288–3297.

[13] D. Liu, K. Song, J.R.G. Townshend, P. Gong, Using local transition probability models in Markov random fields for forest change detection, Remote Sens. Environ. 112 (5) (2008) 2222–2231.

[14] T. Kasetkasem, P.K. Varshney, An image change detection algorithm based on Markov random field models, IEEE Trans. Geosci. Remote Sens. 40 (8) (2002) 1815–-1823.

[15] X. Liu, R.G. Lathrop Jr., Urban change detection based on an artificial neural network, Int. J. Remote Sens. 23 (12) (2002) 2513–2518.

[16] A. Ghosh, N.S. Mishra, S. Ghosh, Fuzzy clustering algorithms for unsupervised change detection in remote sensing images, Inf. Sci. 181 (4) (2011) 699–-715.

[17] G. Pajares, A Hopfield neural network for image change detection, IEEE Trans. Neural Networks 17 (5) (2006) 1250–1264.

[18] J.R. Jensen, Introductory Digital Image Processing: A Remote Sensing Perspective, Prentice Hall, New Jersey, 2005.

[19] N.S. Mishra, S. Ghosh, A. Ghosh, Fuzzy clustering algorithms incorporating local information for change detection in remotely sensed images, Appl. Soft Comput. 12 (2012) 2683–-2692.

[20] S. Ghosh, M. Roy, Modified self-Organizing feature map neural network with semi-supervision for change detection in remotely sensed images, in: S.O. Kuznetsov, D.P. Mandal, M.K. Kundu, S.K. Pal (Eds.), In the Proceedings of Pattern Recognition and Machine Intelligence – 4th International Conference, P R eM I 2011, Moscow, Russia, June 27–July 1, 2011, Lecture Notes in Computer Science, vol. 6744, Springer, Heidelberg, 2011, pp. 98–103.

[21] X. Zhu, Semi-supervised Learning Literature Survey, Computer Sciences TR1530, University of Wisconsin, Madison, 2008.

[22] O. Chapelle, B. Schölkopf, A. Zien, Semi-supervised Learning, MIT Press, Cambridge, 2006.

[23] S. Basu, M. Bilenko, R.J. Mooney, Comparing and unifying search-based and similarity-based approaches to semi-supervised clustering, in: Proceedings of the ICML-2003 Workshop on the Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining Systems, Washington, DC, USA, 2003, pp. 42–49.

[24] K. Wagstaff, C. Cardie, S. Rogers, S. Schroedl, Constrained K-means clustering with background knowledge, in: C.E. Brodley, A.P. Danyluk (Eds.), Proceedings 18th International Conference on Machine Learning (ICML-2001), Williams College/Morgan Kaufmann Publishers Inc., Williamstown, MA, USA/San Fransisco, USA, 2001, pp. 577–584.

[25] D.-Y. Yeung, H. Chang, A kernel approach for semisupervised metric learning, IEEE Trans. Neural Networks 18 (1) (2007) 141–149.

[26] C. Hou, F. Nie, F. Wang, C. Zhang, Y. Wu, Semisupervised learning using negative labels, IEEE Trans. Neural Networks 22 (3) (2011) 420–432.

[27] F. Roli, Semi-supervised multiple classifier systems: background and research directions, in: N.C. Oza, R. Polikar, J. Kittler, F. Roli (Eds.), Proceedings Multiple Classifier Systems, 6th International Workshop, MCS 2005, Lecture Notes in Computer Science, vol. 3541, Springer, Seaside, CA, USA, 2005, pp. 1–11.

[28] Z.-H. Zhou, When semi-supervised learning meets ensemble learning, in: J.A. Benediktsson, J. Kittler, F. Roli (Eds.), Proceedings Multiple Classifier Systems, 8th International Workshop, MCS 2009, Lecture Notes in Computer Science, vol. 5519, Springer, Reykjavik, Iceland, 2009, pp. 529–538.

[29] A. Blum, T. Mitchell, Combining labeled and unlabeled data with co-training, in: P.L. Bartlett, Y. Mansour (Eds.), Proceedings of the 11th Annual Conference on Computational Learning Theory, COLT 1998, Springer, Madison, Wisconsin, USA, 1998, pp. 92–-100.

[30] Z.-H. Zhou, M. Li, Tri-training: exploiting unlabeled data using three classifiers, IEEE Trans. Knowl. Data Eng. 17 (11) (2005) 1529–1541.

[31] S. Patra, S. Ghosh, A. Ghosh, Change detection of remote sensing images with semi-supervised multilayer perceptron, Fundam. Inform. 84 (2008) 429–442.

[32] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice-Hall of India, New Delhi, 2007.

[33] R. Liu, Y. Shi, A RBF classifier with supervised center selection and weighted norm, in: Workshops Proceedings of the 6th IEEE International Conference on Data Mining (ICDM 2006), IEEE Computer Society, Hong Kong, China, 2006, pp. 868–872.

[34] M.-W. Mak, S.-Y. Kung, Estimation of elliptical basis function parameters by the EM Algorithm with application to speaker verification, IEEE Trans. Neural Networks 11 (4) (2000) 961–969.

[35] J.M. Keller, M.R. Gray, J.A. Givens, A fuzzy K-nearest neighbor algorithm, IEEE Trans. Syst. Man Cybern. SMC-15 (4) (1985) 580–585.

[36] I. Partalas, G. Tsoumakas, E.V. Hatzikos, I. Vlahavas, Greedy regression ensemble selection: theory and an application to water quality prediction, Inf. Sci. 178 (20) (2008) 3867–3879.

[37] O. Mendoza, P. Melin, G. Licea, A hybrid approach for image recognition combining type-2 fuzzy logic, modular neural networks and the Sugeno integral, Inf. Sci. 179 (13) (2009) 2078–2101.

[38] R. Xia, C. Zong, S. Li, Ensemble of feature sets and classification algorithms for sentiment classification, Inf. Sci. 181 (6) (2011) 1138–1152.

[39] P. Melin, J. Soto, O. Castillo, J. Soria, A new approach for time series prediction using ensembles of ANFIS models, Expert Syst. Appl. 39 (3) (2012) 3494–3506.

[40] O. Castillo, P. Melin, E. Ramírez, J. Soria, Hybrid intelligent system for cardiac arrhythmia classification with Fuzzy K-Nearest Neighbors and neural networks combined with a fuzzy system, Expert Syst. Appl. 39 (3) (2012) 2947–2955.

[41] J. Li, P.R. Marpu, A. Plaza, J.M.B. Dias, J.A. Benediktsson, A new multiple classifier system for semi-supervised analysis of hyperspectral images., in: The Proceedings of International Conference on Pattern Recognition Applications and Methods (ICPRAM 2012), SciTePress, 2012, pp. 406–411.

[42] B. Waske, J.A. Benediktsson, Semi-supervised classifier ensembles for classifying remote sensing data, in: IEEE International Geoscience and Remote Sensing Symposium, 2008 (IGARSS 2008), vol. 2, 2008, pp. II–105–II–108.

[43] K. Chen, Z. Li, J. Cheng, Z. Zhou, H. Lu, A variational co-training framework for remote sensing image segmentation, in: IEEE International Geoscience and Remote Sensing Symposium, 2009 (IGARSS 2009), vol. 4, 2009, pp. IV–113–IV–116.

[44] M. Li, Z.-H. Zhou, Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples, IEEE Trans. Syst. Man Cybern. – Part A: Syst. Humans 37 (6) (2007) 1088–-1098.

[45] L.I. Kuncheva, J.C. Bezdek, R.P. Duin, Decision templates for multiple classifier fusion: an experimental comparison, Pattern Recognit. 34 (2) (2001) 299–314.

[46] L.I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, John Wiley & Sons Inc., Hoboken, New Jersey, 2004.

[47] A. Halder, A. Ghosh, S. Ghosh, Aggregation pheromone density based pattern classification, Fundam. Inform. 92 (4) (2009) 345–362.

[48] R.G. Congalton, K. Green, Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, 2nd ed., CRC Press, Taylor & Francis Group, Boca Raton, 2009.