Engineering Science and Technology, an International Journal 23 (2020) 973–981

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

journal homepage: www.elsevier.com/locate/jestch

Full Length Article

A novel hybrid machine learning approach for change detection in remote sensing images

Chinmayee Pati [a], Ashok K. Panda [a], Ajaya Kumar Tripathy [b,*], Sateesh K. Pradhan [a], Srikanta Patnaik [c]

[a] Department of Computer Science, Utkal University, Bhubaneswar, India
[b] Department of CSE, Silicon Institute of Technology, Bhubaneswar, India
[c] Department of CSE, SOA University, Bhubaneswar, India

ARTICLE INFO

Article history:
Received 5 August 2019
Revised 7 December 2019
Accepted 11 January 2020
Available online 31 January 2020

Keywords:
Machine learning
Supervised learning
Unsupervised learning
Remote sensing images

ABSTRACT

Change detection can play an essential role in satellite surveillance. With the availability of satellite images of a certain geographical area captured in different time instances, change detection is considered a tough task in the field of satellite applications. This research proposes a novel hybrid machine learning change detection technique from satellite images. The proposed hybrid learning approach is designed based on supervised and unsupervised learning techniques that considers the local association of adjacent pixels of the satellite images. Hybridization of clustering, soft labeling using fuzzy logic, Support Vector Machine (SVM) and Genetic Algorithm (GA) are used in change detection. Radial Basis Function (RBF) is used as the kernel function in SVM, and the RBF kernel parameters such as C and σ are optimized using GA for additional improvement of the performance. To demonstrate the efficiency of the approach, tests are performed on two satellite images captured in two different time instances on a particular geographical area. Change detection accuracy is used to validate the performance. Outcomes are compared with existing approaches and found to be superior.

1. Introduction

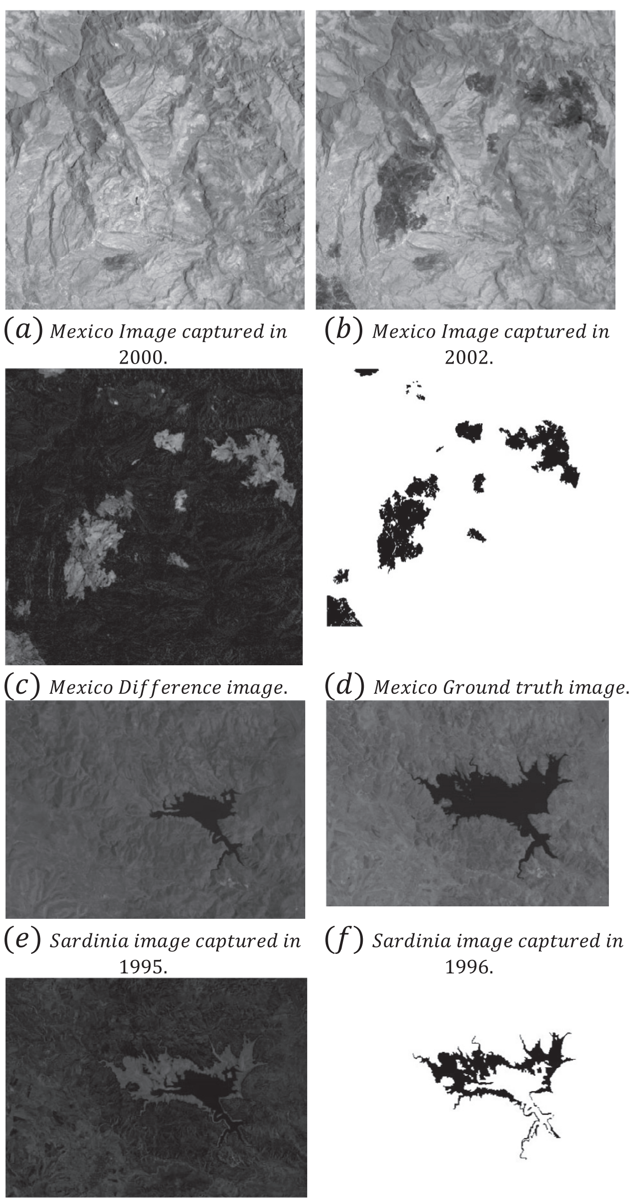
Image analysis, classification and change detection plays a vital role in remote sensing [1,27]. Change detection (CD) in satellite images can play a key role in many different applications [1]. This CD process can be used to identify changes in land covers by the process of remotely sensed multitemporal image segmentation. Some of the reasons for change may be due to the environmental and meteorological incidents such as earthquakes, volcanic eruptions, wildfires, cyclonic storms, floods, drought, heavy snowfall and coastal erosion [2]. CD is considered as one of the most challenging topics in the field of remote sensing application areas [3,4]. CD has vital importance in various applications such as study of land cover dynamics, surveillance of shifting cultivation, forest burning, deforestation, and vegetation change [5]. These problems involve an inspection of larger geographical areas; design of an automatic CD technique is essential to minimize the manual effort and time involved in these applications. CD is employed to identify change in a geographical area by synthesizing multiple images taken at different time instances [6,7]. The major steps involves in the CD process are i) Image preprocessing, ii) finding difference image (DI), and iii) difference image analysis. We use the change vector analysis (CVA) technique to produce DI from images of the same geographical location at two different times. Each pixel of the DI represents a fixed amount of geographical area. The task is to categorize the set of pixels in the DI into two categories such as changed or unchanged groups. This process of categorization of pixels can be carried out using a supervised as well as an unsupervised approach [8–12].

Several unsupervised techniques were proposed based on change vector analysis by applying a thresholding approach on the magnitude of the DI pixels [1,13,14]. The limitation of this approach is that the magnitude operator loses some information about the direction of difference. Some of the studies apply a clustering technique to solve the problem but in many real applications the accuracy is lower than those provided by thresholding methods [15]. Again, clustering approaches need an additional physical post-processing for labeling the clusters with changed and unchanged labels which limits the automation of the approach. Hence, several other effective unsupervised approaches have been designed in the literature [16,17,24–26]. Although, the supervised approach is performing better than the unsupervised approach, it requires large amount of labeled data which is not available in reality.

* Corresponding author.
E-mail address: ajayatripathy1@gmail.com (A.K. Tripathy).
Peer review under responsibility of Karabuk University.

https://doi.org/10.1016/j.jestch.2020.01.002
2215-0986/© 2020 Karabuk University. Publishing services by Elsevier B.V.

This research, tackles the aforesaid problem by designing a hybrid technique for change detection in satellite images using optimized SVM and Fuzzy c-means. The objective of the planned technique is to use essential unsupervised technique to obtain a Fuzzy set for changed pixels with degree of change as Fuzzy membership function. Further it applies α-cut to auto label the dataset with a suitable α-value. Then, it applies supervised techniques with the auto labeled data and pre-collected small amount of ground reality data (through sensors or through manual surveillance). Support Vector Machine (SVM) is one of the most effective approaches among supervised techniques [16,17]. The accuracy of SVM depends on suitable kernel and kernel parameter setting. Under this situation, instead of using heuristic based kernel and kernel parameter settings, a GA optimization technique can be applied to select the best suitable kernel and kernel parameter settings. SVM optimization using GA can be found in the literature [18]. This approach uses GA optimized SVM (O-SVM) classifier to handle the supervised learning.



(a) Mexico Image captured in 2000. (b) Mexico Image captured in 2002. (c) Mexico Difference image. (d) Mexico Ground truth image. (e) Sardinia image captured in 1995. (f) Sardinia image captured in 1996. (g) Sardinia difference image (h) Sardinia Ground truth image.

Fig. 1. Satellite pictures of Mexico-area and Sardinia Island-area of Italy.

The rest of the paper is organized as follows. Section 2, describes the dataset used in this paper. The feature extraction process is explained in Section 3. A hybrid machine learning approach for CD is proposed in Section 4. A detailed description of the supervised approach used in the work is presented in Section 5. Section 6, presents the experimental evaluation and analysis report. Finally, Section 7 concludes the work presented in this paper.

2. Description of dataset

To assess the performance of the proposed approach, tests are performed on two multi-temporal remotely sensed images analogous to the geographical areas of Mexico and Sardinia Island of Italy. The Mexico area images were captured from the Landsat-7 satellite on a part of the Mexico geographical area on date $18:04:2000$ and on date $20:05:2002$ which are shown in Fig. 1(a) and (b) respectively. From the total scene, a part of 512×512 pixels area is chosen as the investigating area. During these two years, a fire destroyed a big portion of the area under consideration. The difference image is shown in Fig. 1(c). To validate the effectiveness of the approach, a reference map is used called ground truth image, shown in Fig. 1(d). The ground truth image consists of 25599 changed and $236,545$ unchanged pixels. Sardinia Island, Italy images were captured in *Sept* 1995 and *July* 1996 which are shown in Fig. 1(e) and (f) respectively. From the total scene, a part of 412×300 pixels area is selected as the investigation area. The water level of the lake elevated between the two images captured dates. The difference image and ground truth image are shown in Fig. 1(g) and (h) respectively. The ground truth image consists of 7480 changed and 116120 unchanged pixels. The datasets are collected from [5].

Algorithm 1: Spatial feature extraction algorithm

Given: The two images M_1, M_2 captured at two different time instances and radial distance d to be considered for spatial information.
Objective: To generate a dataset with spatial features related to each Pixel Location.
Procedure:
1. $DI[i,j]=|M_1[i,j]-M_2[i,j]|$
2. *for* each row i in DI:
3. *for* each column j in DI:
4. $DS[i \times n + j] =$ All pixel values within the radial distance d, where n is the number of columns in DI. That constitutes the generated features for the pixel position $[i,j]$.
5. The pixel intensity $PI[i \times n + j] = DI[i,j]$
6. *end for*
7. *end for*
8. *return DS and PI*

3. Feature extraction from images

The given remote sensing gray scale images of a particular area are taken at two altered time instances. Each of the images can be considered as an integer matrix each of size $m \times n$ (m: rows and n: columns) with value ranges from 0 to 255. Each element of the matrix represents a pixel of the image. Each pixel of the image represents a particular location of the geographical location from which the image is captured which is called as Pixel Location. Let these two matrices be represented as M_1 and M_2. The objective is to identify the locations where there is a change. This objective can be restated as, identify the pixels which represent the changed location at time instance t_2 with respect to time instance t_1. Spatial information can play a key role in identifying a pixel as changed or unchanged. To consider the spatial information we have collected the pixel values as well as the neighbouring pixel values which lie within a radial distance d for feature extraction. Distance between vertical or horizontal neighbouring pixels is considered as a unit distance. The spatial feature extraction process is explained in Algorithm 1. In case of multiple band images the pixel intensity (PI) of DI is produced by the CVA technique [28] i.e., using Eq. (1).

$$PI_{ij} = (int)\left|\sqrt{\sum_{b=1}^{n}\left(PI_{ij}^{b}(M_1) - PI_{ij}^{b}(M_2)\right)^2}\right| \tag{1}$$

4. Hybrid machine learning approach for change detection

In the present work, a hybrid (combination of unsupervised and supervised) machine learning approach is proposed for CD. The proposed approach has two major parts: in the first part unsupervised approach is applied to learn membership value or belongingness of each pixel in both the classes, and then along with these auto generated class labels, few labeled pixels are used for the proposed classifier learning. Finally, using the learned classifier all the pixels are classified and the performance accuracies are measured. In our experiment, identical quantities of supervised data are selected at random. The proposed CD approach is depicted in Fig. 2. Detailed description of each of the approach is presented in subsequent sections.

Algorithm 2: Automatic Pixel soft-labelling

Given: The dataset DS and the pixel intensity set PI generated by Algorithm 1. DS is defined as $DS = \left\{p_1, p_2, \cdots, p_{np}\right\}$, where $p_i =< p_i^1, p_i^2, \cdots, p_i^f >$ is a f dimensional vector, f:is the number of features and np:is the number of pixels. PI is defined as $PI = \{i_1, i_2, \cdots, i_{np}\}$, where i_k is an integer value which represent each pixel intensity.
Objective: To label the data points to one of the two labels: Changed or Unchanged.
Procedure:
1. Randomly select two data points r_1 *and* r_2 within the range $[p_1,\ p_{np}$ as two representatives.
2. *for* k in range $[1, np]$:
3. $d_1 = \sqrt{\sum_{i=1}^{f}(p_k[i] - r_1[i])^2}$
4. $d_2 = \sqrt{\sum_{i=1}^{f}(p_k[i] - r_2[i])^2}$
5. $p_k^{label} = p_k^l = ArgMin_x(d_x)$
6. *end for*
7. $r_1^{old} = r_1;\ r_2^{old} = r_2;$
8. *for* m in range $[1, f]$:
9. $FS_m^1 = 0; FS_m^2 = 0; c_1 = 0; c_2 = 0;$
10. *for* n in range$[1, np]$:
11. *if* $p_m^l == 1$:
12. $FS_m^1 = FS_m^1 + DS[n, m]; c_1 + +;$
13. *else*
14. $FS_m^2 = FS_m^2 + DS[n, m]; c_2 + +;$
15. *end for*
16. $r_1[m] = \frac{FS_1^m}{c_1};\quad r_2[m] = \frac{FS_2^m}{c_2};$
17. *end for*
18. $E_1 = \sqrt{\sum_{i=1}^{f}(r_1^{old}[i] - r_1[i])^2}$
19. $E_2 = \sqrt{\sum_{i=1}^{f} r_2^{old}[i] - r_2[i]^2}$

(*continued*)

Algorithm 2: Automatic Pixel soft-labelling

20. *if* $(E_1 > \varepsilon or E_2 > \varepsilon)$:
21. *GO TO* Step 2
22. $PIS_1 = 0;\quad PIS_2 = 0;\quad c_1 = 0;\quad c_2 = 0;$
23. *for* i in range$[1, np]$:
24. *if* $p_i^l == 1$:
25. $PIS_1 = PIS_1 + PI[i]; c_1{+}{+};$
26. *else*
27. $PIS_2 = PIS_2 + PI[i]; c_2{+}{+};$
28. *end for*
29. $MPIS_1 = \frac{PIS_1}{c_1};\quad MPIS_2 = \frac{PIS_2}{c_2};$
30. *if* $(MPIS_1, MPIS_2)$:
31. $Label = \{1:\ Changed,\ 2:\ Unchanged\}$
32. *else*
33. $Label = \{1:\ Unchanged,\ 2:\ Changed\}$
34. *return* $p^l =< p_1^l, p_2^l, \cdots, p_{np}^l >$

4.1. Proposed unsupervised change detection approach

Each pixel in the image represents a particular region of geographical location. The objective of the CD problem is to identify the regions where there is a change. Alternatively, the problem is to label the pixels to one of the two labels- Changed or Unchanged. An unsupervised approach has been proposed to make an initial level soft prediction of each pixel class labeling. The pixel soft-labelling process is described in Algorithm 2. Further, using the soft-labeling two fuzzy sets are designed to represent the changed group and the unchanged group. The fuzzy set design process is described in Algorithm 3.

4.2. Training and test dataset for supervised learning

Upon completion of unsupervised learning all pixels are the members of two fuzzy sets namely, Changed (CH) and Unchanged (UC). The fuzzy set CH is defined as

$$CH = \{i/\mu_{CH}(i)\}_{i=1}^{np},\ \text{and the fuzzy set UC is defined as} \tag{1}$$

$$UC = \{i/\mu_{UC}(i)\}_{(i=1)}^{np},\quad where\ np = m \times n. \tag{2}$$

The membership value μ_{CH} and μ_{UC} are the output of Algorithm 3. For the supervised learning we will use few ground truth pixel labels (less than15%) of G and labels from alpha cut of both the fuzzy sets CH and UC. In our experiment, the value of alpha was taken as 0.6, 0.7, 0.8and 0.9. The training set will be $\{G \cup CH \cup UC\}$ and the test set will be $\{DS\}$.

Algorithm 3: Changed and Unchanged fuzzy set design process

Given: The pixel intensity set PI generated by Algorithm 1. PI is defined as $PI = \{i_1, i_2, \cdots, i_{np}\}$, where i_k is an integer value which represents k^{th}pixel-intensity. Each pixel's soft-label $p^l =< p_1^l, p_2^l, \cdots, p_{np}^l >$ and *Labels* generated by Algorithm 2.
Objective: To define two fuzzy sets representing the changed (CH) and unchanged (UC) class with membership function $\mu_{CH}(i)$ *and* $\mu_{UC}(i)$.
Procedure:
1. Let $cl = Label(changed);\quad max = 0,\ min = \infty;$
2. *for* i in range $[1, np]$:
3. *if* $(p_i^l == cl \&\& PI[i] > max)$:

(*continued on next page*)

(continued)

Algorithm 3: Changed and Unchanged fuzzy set design process
4. $max = PI[i]$
5. *else if* $(p_i^l \neq cl \&\& PI[i] < \min)$:
6. $min = PI[i]$
7. *end for*
8. *for i*in range $[1, np]$:
9. *if* $(p_i^l == cl)$:
10. $\mu_{CH}(i) = \frac{PI[i]}{max}$; $\mu_{UC}(i) = 0$;
11. *else*
12. $\mu_{UC}(i) = \frac{PI[i]}{min}$; $\mu_{CH}(i) = 0$;
13. *end for*
14. *return* μ_{CH} *and* μ_{UC}

5. OSVM: proposed supervised approach for change detection

The proposed Optimized Support Vector Machine (OSVM) is a supervised machine learning approach for linearly nonseparable datasets. This approach is designed by optimizing the Support Vector Machine kernel parameters using GA.

5.1. Overview of the SVM classifier

Assumption: The problem in hand is a supervised binary, linearly classifiable problem. The training sample can be defined as $\mathrm{F_i} \in \mathcal{R}^{\mathrm{n}}, \mathrm{i} = 1, 2, \cdots, \mathrm{M}$. i.e, the training sample set consists of M vectors in N dimensional feature space. Each training sample X_i is associated with a class level $l_i \in \{+1, -1\}$.

Since it is assumed that the classes are linearly separable, there exists at least one hyperplane which can separate both the classes. Let the equation of that hyperplane is $Z^T F + a = 0$, where F is a feature vector, $Z \in \mathcal{R}^n$ is a flexible weight matrix, and bias $a \in \mathcal{R}$.

$$\therefore Z^T F_i + a \geq 0 \quad if\ l_i = +1; \tag{3}$$

$$and\ Z^T F_i + a \leq 0 \quad if\ l_i = -1; \tag{4}$$

There may be infinite number of hyperplanes that separate the two classes but we have to find out a hyperplane for which the separation margin from the closest data points is maximum. This hyperplane is called optimal hyperplane for separating the two classes.

The optimal Z and a can be obtained through the following optimization problem (p):

Given the training sample $\{(F_i, l_i)\}_{i=1}^{M}$, find Z and a such that $l_i\left(Z^T F_i + a\right) \geq 1,\ for\ i = 1, 2, \cdots, M$ and Z minimizes $\varphi(Z) = \frac{1}{2}||Z||^2$.

This problem can be treated as a primal problem, as $\varphi(Z)$is a convex function of Z and the constraints are linear in Z.

The constraint optimization problem can be solved using Lagrange multipliers. We can construct the Lagrange formulation as:

$$f(z, \alpha, \beta) = \frac{1}{2}||Z||^2 - \sum_{i=1}^{M} \beta_i \left[t_i \left(Z^T F_i + a \right) - 1 \right] \tag{5}$$

where $\beta_i \geq 0$ are called Lagrange multiplier. The solution to this optimization problem requires that

$$\frac{\partial f(Z, \alpha, \beta)}{\partial Z} = 0 \tag{6}$$

$$\frac{\partial f(Z, \alpha, \beta)}{\partial \alpha} = 0 \tag{7}$$

By applying Eq. (6), Eq. (5) yields

$$Z = \sum_{i=1}^{M} \beta_i t_i F_i \tag{8}$$

By applying Eq. (7), Eq. (5) yields

$$\sum_{i=1}^{M} \beta_i t_i = 0 \tag{9}$$

Putting Eqs. (8) and (9) into Eq. (5), the problem is converted to

$$f(\beta) = \sum_{i=1}^{M} \beta_i - \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \beta_i \beta_j l_i l_j F_i^T F_j, \quad where\ \beta_i \geq 0 \tag{10}$$

Now the dual problem can be formulated as: Given the training sample $\{(F_i, t_i)\}_{(i=1)}^{M}$, find the Lagrange Multipliers $\{\beta_i\}_{i=1}^{M}$ that maximize the objective function

$$f(\beta) = \sum_{i=1}^{M} \beta_i - \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \beta_i \beta_j l_i l_j F_i^T F_j, \quad where\ \beta_i \geq 0 \tag{11}$$

with constraints $\sum_{i=1}^{M} \beta_i l_i = 0\ and\ \beta_i \geq 0, \quad i = 1, 2, \cdots, M$.

The optimal hyperplane can be designed by maximizing $f(\beta)\ w.r.to\ \beta_i \geq 0$ subject to the constraints $\sum_{i=1}^{M} \beta_i l_i = 0\ and\ \beta_i \geq 0, \quad i = 1, 2, \cdots, M$.

5.2. Non-Linear SVM

In this approach the training samples are mapped into higher-dimensional feature space through a proper nonlinear transformation function φ into a higher dimensional space $\varphi(F) \in \mathcal{R}^{N'}, N' > N$. In this $\mathcal{R}^{N'}$ the separation between the two classes can be looked as a linearly separable case. Therefore, the optimization problem defined in Eq. (11) can be rewritten by replacing (F_i, F_j) with inner product in the transformed space $(\varphi(F_i), \varphi(F_j))$. But the main problem is to compute $\varphi(F)$, which can prove to be expensive and time consuming. However, the kernel method provides an alternative to deal with this problem. Let us assume that a kernel function satisfies the above stated transformation condition i.e. $g(F_j, F_i) = \varphi(F_j), \varphi(F_i)$. Therefore, the optimized problem for a non-linear case can be formulated as: with $\{F_i, t_i\}_{i=1}^{M}$, find $\{\beta_i\}_{i=1}^{M}$ that maximizes

$$f(\beta) = \sum_{i=1}^{M} \beta_i - \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \beta_i \beta_j l_i l_j g(F_i, F_j) \tag{12}$$

with constraints $\sum_{i=1}^{M} \beta_i l_i = 0\ and\ o \leq \beta_i \leq r, \quad i = 1, 2, \cdots, M$.where, r is the regularization parameter.

The discriminate function $D(F)$ can be expressed as:

$$D(F) = \sum_{i=1}^{M} \beta_i l_i \mathcal{g}(F_i, F) + \alpha \tag{13}$$

Kernel-based SVM improves the performance of SVM in most of the real life data sets, but it involves the problem of multiple parameter selection such as kernel parameters and regularization parameter r. This paper aims to automate the process of parameter selection for Kernel-based SVM using Genetic Algorithm.

Genetic Algorithm: Genetic Algorithm (GA) tackles the problem of searching for the best hypothesis from a space of candidate hypothesis. The best hypothesis is defined as the one that optimizes a predefined fitness measure. GA works by an iterative updating approach on a set of hypothesis called initial population. The fitness value of each hypothesis is tested in each iteration. A new set of hypothesis is then generated by choosing some of the best-fit hypothesis from the current set of hypothesis and by producing new offspring individuals through genetic operations. These new set of hypothesis are carried forwarded to further iterations.

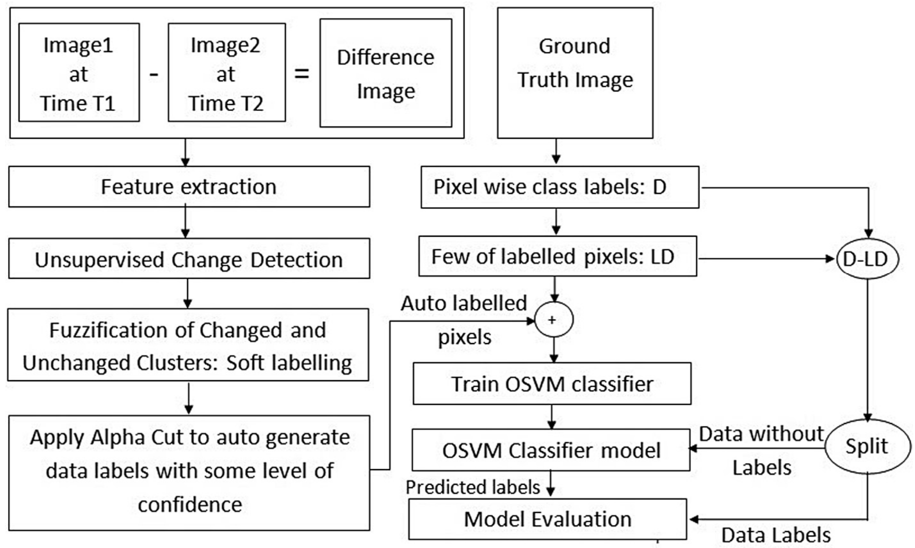


Fig. 2. Architecture of the proposed change detection approach.

The fitness function tests the fitness of a hypothesis in each generation. The genetic operations such as crossover and mutation operations are the prime functions that arbitrarily influence the quality. Hypotheses are chosen for regeneration through their quality. The high quality ones having higher chances are chosen for the crossover pool using roulette wheel or the tournament selection process. The evolutionary process continues until the termination condition is satisfied.

5.3. GA-based SVM parameter regularization

Instead of using hypothetical kernel parameters for SVM, GA works with a set of candidate solutions called hypothesis space. To demonstrate the approach, a RBF kernel for SVM classifier is selected. The RBF kernel needs two parameters, C and γ which must be optimized. This work proposes a GA based RBF kernel parameter optimization for CD. The GA chromosome comprises of two parts, C and γ. But, these chromosomes have changed factors for different kernels. Binary encoding is used to represent the chromosome and CD accuracy is used for fitness function. The structure of chromosome is $\left(b_c^p, b_c^{p-1}, \cdots, b_c^1, b_\gamma^q, b_\gamma^{q-1}, \cdots, b_\gamma^{q1}\right)$ of $p+q$ bits, where b represents one bit. p is the number of bits representing parameter C, q is the number of bits representing parameter p and q can be chosen according to the calculation precision as required. The precision of kernel parameters depend on values of p and q; and the permissible minimum and maximum value of the parameter is predefined. Value of the parameter is decided by Eq. (14).

$$\alpha_{val} = \frac{[\alpha_{max} - \alpha_{min}]}{2^k - 1} \times \sum_{i=1}^{k} b_\alpha^i \times 2^{i-1} \tag{14}$$

where, α_{val} is the value of parameter α in a chromosome, α_{max} & α_{min} are the maximum and minimum permissible values of α, k is the number of bits used for encoding α, b_α^i is the i^{th} bit value of α inside a chromosome.

Fitness Function: Pixel labeling accuracy percentage is the fitness function. Thus, a hypothesis with higher accuracy produces a high fitness value. The hypothesis with high fitness value has high probability to be passed onto the next generation.

Stopping Criteria: The process of generating new generation of hypothesis by applying genetic operators will stop after the accuracy reaches near 100%, or if there is no improvement in accuracy for last few generations, or the maximum number of iterations (500) are reached.

Parameter Setting for GA: Population size 200, crossover rate 0:75, mutation rate 0:01, two point crossover, and Roulette wheel selection.

6. Experimental results and analysis

To tune the hyper parameters, experiments are performed on remote-sensing images. The proposed work consists of three hyper parameters: 1) radial distance d for feature generation, 2) Value of α for $\alpha - cut$, and 3) % of ground truth data β. To obtain the optimum value of these parameters, we have performed experiments on combination of values for these three parameters. For experimentation, one radial distances for feature generation $\left(d = \sqrt{2}\right)$, three different $\alpha - cut$ used for soft labeling $(\alpha \in \{0.6, 0.7, 0.8\})$, and three different percentage of ground truth data $(\beta \in \{2, 4, 5\})$ are considered. For validation, 50-fold, 25-fold and 20-fold cross validations are conducted for 2%, 4% and 5% training samples respectively. The experiments are coded with two symbols $[\alpha, \beta]$. Two type of errors are considered in the prediction of labels: 1) False Positive (FP), where pixel ground truth label is unchanged but the predicted label is changed, 2) False Negative (FN), where pixel ground truth label is changed but the predicted label is unchanged. Total error, $E = FP + FN$. In a $k - fold$ experiment the average error

$$AE = \frac{\sum_{i=1}^{k} E_i}{k}, \quad \text{is taken}$$

The results obtained on both the datasets for $d = 1$ *and* 1.414 are shown in Table 1. These results indicate that the hyper parameters $d = 1.414$ and $\alpha = 0.7$ perform better when $\beta = 5$. It can also be observed that by increasing β from 2 to 4 there is a significant decrease in error whereas, by increasing β from 4 to 5 the error reduces negligibly. Since, in reality the availability of ground truth data is small in nature, we have selected $\beta = 4$ for further experiment with $d = 1:414$ and $\alpha = 0.7$. For visual comparison of the hyper parameters the bar plots of the results are plotted in Fig. 3 and 4 for the Mexico and Sardinia datasets respectively.

Table 1
Hyper parameters (α,β and d) tuning experiment results.

Experiment details	αValue for $\alpha - cut$	% of Supervised data used	K-fold Average False Positive	K-fold Average False Negative	K-fold Average Total Error
Mexico data with d = 1	0.6	2	2446	537	2983
		4	2179	652	2831
		5	2085	589	2674
	0.7	2	1899	506	2705
		4	1875	732	2607
		5	2122	465	2587
	0.8	2	2179	766	2945
		4	2297	611	2908
		5	2086	696	2782
Mexico data with $d = 1.414$	0.6	2	2344	589	2933
		4	2106	703	2809
		5	2073	552	2625
	0.7	2	1991	630	2621
		4	2029	508	2537
		5	2083	427	2510
	0.8	2	2532	413	2945
		4	2179	689	2868
		5	2085	623	2708
Sardinia data With $d = 1$	0.6	2	1144	286	1430
		4	958	287	1245
		5	819	273	1092
	0.7	2	879	278	1157
		4	838	184	1022
		5	867	166	1033
	0.8	2	1280	209	1489
		4	1064	218	1282
		5	921	307	1228
Sardinia data With $d = 1.414$	0.6	2	1125	282	1407
		4	906	302	1208
		5	793	238	1031
	0.7	2	926	211	1137
		4	838	175	1013
		5	773	232	1005
	0.8	2	1189	264	1453
		4	1037	212	1249
		5	967	211	1178

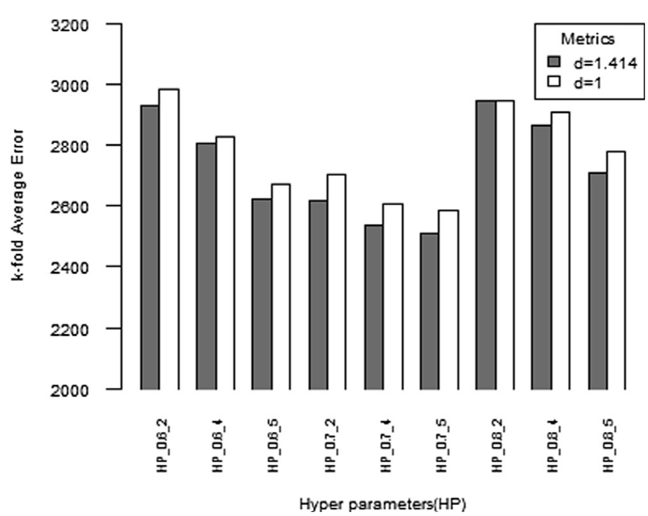


Fig. 3. Results for different Hyper parameters HP_α_β and d for MexicoData.

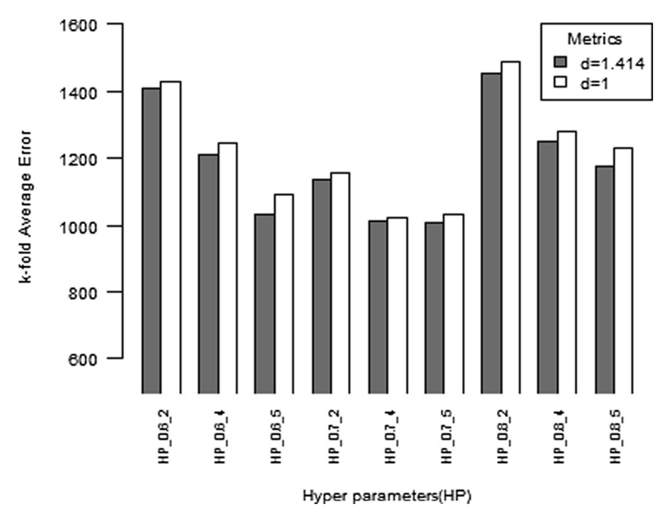


Fig. 4. Results for different Hyper parameters HP_α_β and d for Sardinia Data.

As the proposed approach is a hybrid approach which is a combination of supervised and unsupervised machine learning, to choose suitable techniques to be used for both the components of the approach, several experiments are performed. The unsupervised technique such as density based clustering technique Revised DBSCAN [23], k-means [21], and Fuzzy c-Means (FCM) [22] are considered. SVM, OSVM with soft labeling, RBF network (15 hidden unit: 9-15-1), MLP1 (Shallow network: structure-9-20-1, activation function used- sigmoid), MLP2 (Deep network: structure-9-5-5-5-1, activation function used- sigmoid) are considered as supervised approaches. Experiments are carried out by considering one clustering and one classification technique at a time. Five % of the training patterns are used for supervised learning and 20-fold cross validation are performed. The results obtained for both datasets are presented in Table 2. For visual comparison, both of the bar plots are presented in Fig. 5 and 6. The results shows that the

Table 2
Supervised and unsupervised machine learning model selection.

Dataset Used	Approaches Used	20-fold Average MA	20-fold Average FA	20-fold Average TE
Mexico dataset	DBSCAN – SVM	2273	719	2992
	DBSCAN – OSVM with soft labeling	2021	646	2677
	DBSCAN – RBF NW	2799	894	3693
	DBSCAN –MLP1	2795	883	3678
	DBSCAN –MLP2	2635	831	3466
	k-Means – SVM	2484	783	3267
	k-Means – OSVM with soft labeling	2120	671	2791
	k-Means – RBF NW	2817	893	3710
	k-Means –MLP1	2911	919	3830
	k-Means –MLP2	2813	892	3705
	FCM – SVM	2185	694	2879
	FCM – OSVM with soft labeling	1907	603	2510
	FCM – RBF NW	2783	882	3665
	FCM – MLP1	2729	859	3588
	FCM – MLP2	2881	906	3787
Sardinia dataset	DBSCAN – SVM	931	280	1211
	DBSCAN – OSVM with soft labeling	834	254	1097
	DBSCAN – RBF NW	1025	308	1333
	DBSCAN –MLP1	1038	312	1350
	DBSCAN –MLP2	993	295	1288
	k-Means – SVM	963	290	1253
	k-Means – OSVM with soft labeling	875	263	1138
	k-Means – RBF NW	1083	322	1405
	k-Means –MLP1	1125	328	1453
	k-Means –MLP2	1073	326	1399
	FCM – SVM	902	271	1173
	FCM – OSVM with soft labeling	773	232	1005
	FCM – RBF NW	1033	312	1345
	FCM – MLP1	1012	305	1317
	FCM – MLP2	1059	318	1377

FCM + OSVM with soft labeling provides batter results compared to other combinations.

To assess the efficiency of the proposed technique, tests are conducted on both of the datasets and the performance recorded using the proposed approach are compared with three supervised methods based on SVM, MLP [19] and RBF network [20], and two unsupervised approaches based on $k-means$ [21] and $Fuzzyc-means$ [22]. For the proposed approach, 5% of training patterns are considered and 20-fold cross validations are performed to validate the accuracy. For supervised approaches 70% training data and 30% testing data are used. For SVM approach the value of $c = 1$ and $\gamma = 0.2$. For MLP the network has 9 input units in input layer, 2 inner layers with 3 units per layer and 1 node in output layer. The activation function used in each unit is Sigmoid. The RBF network has 9 nodes in input layer and 20 nodes in hidden layer and 1 node in output layer. The results obtained for both the datasets are presented in Table 3. For visual comparison a barplot is presented in Fig. 7. These results show that the supervised approaches provide better results compared to the unsupervised approaches. Whereas, in reality getting large number of labeled data is not possible in this problem. However, the proposed hybrid approach gives a superior results with few (only 5%) labeled data.

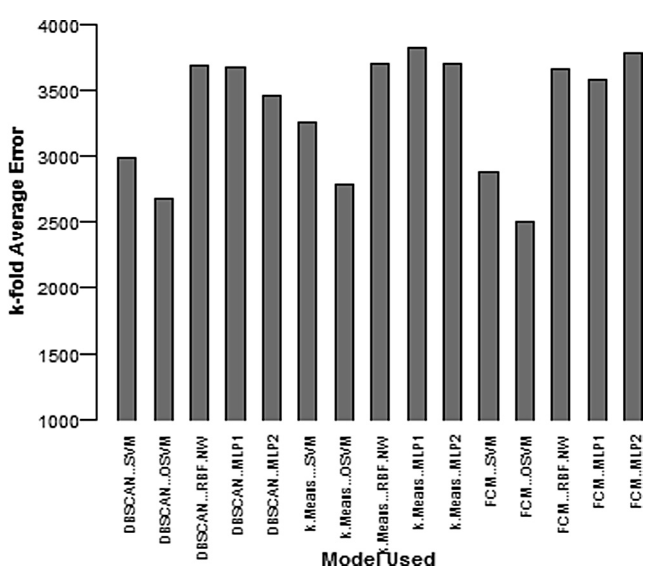


Fig. 5. Supervised and unsupervised machine learning model selection: Mexico Data.

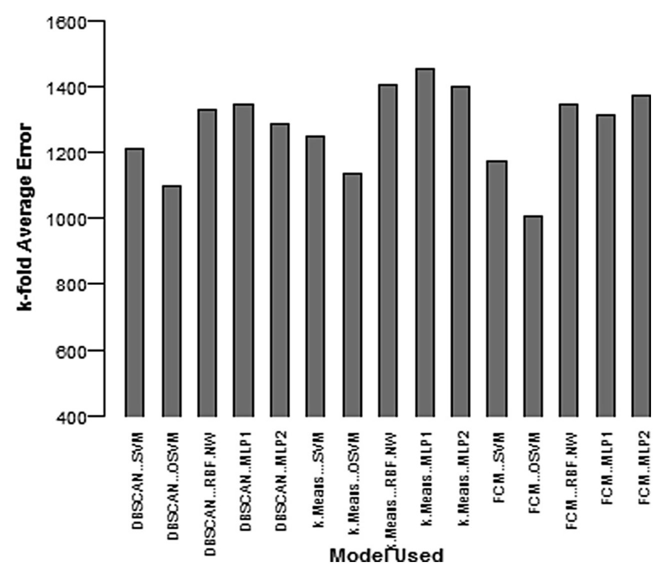


Fig. 6. Supervised and unsupervised machine learning model selection:Sariania Data.

Table 3
Performance comparison by applying supervised and unsupervised learning model independently.

Dataset Used	Approaches Used	20-fold Average MA	20-fold Average FA	20-fold Average TE
Mexico dataset	Supervised SVM	1262	316	1578
	Supervised MLP	1380	305	1685
	Supervised RBF Network	1325	340	1665
	Unsupervised$2-Means$	3113	879	3992
	Unsupervised$Fuzzy2-Means$	2835	945	3780
	PA	1907	603	2510
Sardinia dataset	Supervised SVM	705	202	907
	Supervised MLP	790	198	988
	Supervised RBF Network	685	294	979
	Unsupervised$2-Means$	1312	393	1705
	Unsupervised$Fuzzy2-Means$	1269	401	1670
	PA	773	232	1005

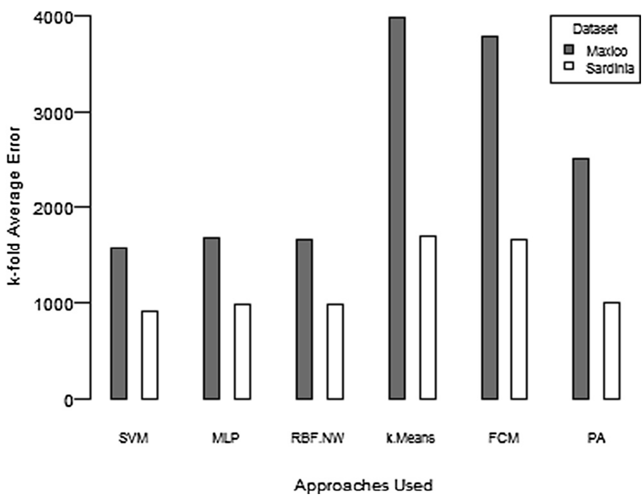


Fig. 7. Supervised and unsupervised learning model independently.

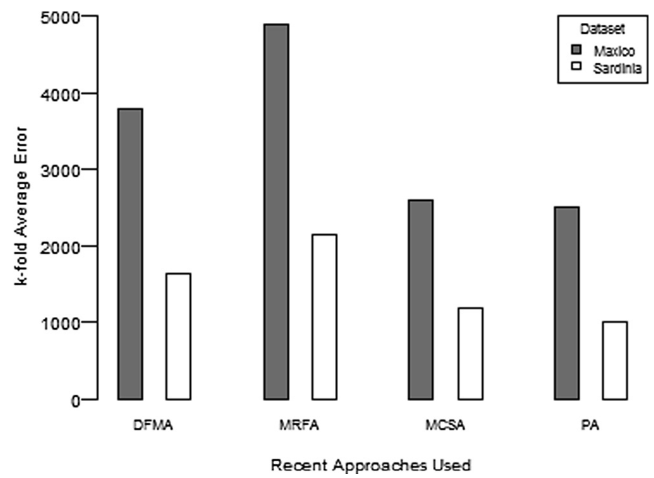


Fig. 8. Performance comparison with recent approaches.

Table 4
Performance comparison with recent approaches.

Dataset Used	Approaches Used	20-fold Average MA	20-fold Average FA	20-fold Average TE
Mexico dataset	DFMA	2871	912	3783
	MRFA	3725	1172	4897
	MCSA	1971	664	2605
	PA	1907	603	2510
Sardinia dataset	DFMA	1251	378	1629
	MRFA	1653	504	2157
	MCSA	901	272	1173
	PA	773	232	1005

The proposed approach is compared with the recent approaches. The change detection results of a Markov Random Field based approach (MRFA) [24], Deep Feature Mapping based Approach (DFMA) [25], and Multiple classifier System based Approach (MCSA) [26] are compared with the Proposed Approach (PA). Five percent of training patterns are used for supervised learning and 20-fold cross validation are performed. The results obtained for both datasets are presented in Table 4. For visual comparison, a bar plot is presented in Fig. 8. The results show that the proposed approach provides better results compared to the recent approaches.

7. Conclusion

In this paper, a novel approach for change detection is proposed using hybrid machine learning. In this approach a soft labeling is generated using unsupervised learning. Further, along with the soft labeling and little ground truth data a supervised learning is applied using OSVM. In the present investigation, $fuzzy\ c-means$ is used as base method for unsupervised learning. Experiments are performed on multitemporal satellite datasets to validate the effectiveness of the proposed approach. From the results, it is observed that the proposed hybrid machine learning approach outperforms the other state-of-the-art approaches when a small amount of supervised data is available.

Acknowledgments

The author would like to thank Prof. Jaideep Talukdar for his thorough and constructive comments which helped to enhance the quality of this paper.

References

[1] A. Singh, Digital change detection techniques using remotely-sensed data, Int. J. Remote Sens. 10 (6) (1989) 989–1003.
[2] A.M. Roy, S. Ghosh, A. Ghosh, A neural approach under active learning mode for change detection in remotely sensed images, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 (4) (2014) 1200–1206.
[3] M. Zanetti, L. Bruzzone, A generalized statistical model for binary change detection in multispectral images, in: Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International, IEEE, 2016, pp. 3378–3381.
[4] A.P. Tewkesbury, A.J. Comber, N.J. Tate, A. Lamb, P.F. Fisher, A critical synthesis of remotely sensed optical image change detection techniques, Remote Sens. Environ. 160 (2015) 1–14.
[5] 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.

[6] L. Bruzzone, D.F. Prieto, An mrf approach to unsupervised change detection, in: Proceedings of International Conference on Image Processing: ICIP 99, IEEE, 1999, pp. 143–147.
[7] K. Islam, M. Jashimuddin, B. Nath, T.K. Nath, Land use classification and change detection by using multi-temporal remotely sensed imagery: the case of chunati wildlife sanctuary, bangladesh, Egypt. J. Remote Sens. Space Sci. 21 (1) (2018) 37–47.
[8] F. Yuan, K.E. Sawaya, B.C. Loeffelholz, M.E. Bauer, Land cover classification and change analysis of the twin cities (minnesota) metropolitan area by multitemporal landsat remote sensing, Remote Sens. Environ. 98 (2) (2005) 317–328.
[9] G. Camps-Valls, L. Gómez-Chova, J. Munoz-Marı, J.L. Rojo-Alvarez, M. Martınez-Ramon, 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] F. Melgani, G. Moser, S.B. Serpico, Unsupervised change-detection methods for remote-sensing images, Opt. Eng. 41 (12) (2002) 3288–3297.
[11] D. Liu, K. Song, J.R. Townshend, P. Gong, Using local transition probability models in markov random fields for forest change detection, Remote Sens. Environ. 112 (5) (2008) 2222–2231.
[12] G. Pajares, A hopfield neural network for image change detection, IEEE Trans. Neural Networks 17 (5) (2006) 1250–1264.
[13] L. Bruzzone, D.F. Prieto, Automatic analysis of the difference image for unsupervised change detection, IEEE Trans. Geosci. Remote Sens. 38 (3) (2000) 1171–1182.
[14] F. Bovolo, L. Bruzzone, A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain, IEEE Trans. Geosci. Remote Sens. 45 (1) (2007) 218–236.
[15] F. Bovolo, L. Bruzzone, M. Marconcini, A novel approach to unsupervised change detection based on a semisupervised svm and a similarity measure, IEEE Trans. Geosci. Remote Sens. 46 (7) (2008) 2070–2082.
[16] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[17] E. Tuba, M. Tuba, M. Beko, in: Support Vector Machine Parameters Optimization by Enhanced Fireworks Algorithm, Springer, 2016, pp. 526–534.
[18] C.-L. Huang, C.-J. Wang, A ga-based feature selection and parameters optimizationfor support vector machines, Expert Syst. Appl. 31 (2) (2006) 231–240.
[19] M.T. Hagan, H.B. Demuth, M.H. Beale, O. De Jesus, Neural Network Design, Pws Pub, Boston, 1996.
[20] M.J. Orr, et al., Introduction to radial basis function networks, 1996.
[21] S. Basu, A. Banerjee, R. Mooney, Semi-supervised clustering by seeding, Proceedings of 19th International Conference on Machine Learning (ICML-2002. Citeseer), 2002.
[22] N.S. Mishra, S. Ghosh, A. Ghosh, Fuzzy clustering algorithms incorporating local information for change detection in remotely sensed images, Appl. Soft Comput. 12 (8) (2012) 2683–2692.
[23] T.N. Tran, K. Drab, M. Daszykowski, Revised DBSCAN algorithm to cluster data with dense adjacent clusters, Chemometr. Intell. Lab. Syst. 120 (2013) 92–96.
[24] A.G. Wei, L. Zhihan, H. Ming, Change detection method for remote sensing images based on an improved Markov random field, Multimedia Tools Appl. 76 (17) (2017) 17719–17734.
[25] P. Zhang, M. Gong, L. Su, J. Liu, Z. Li, Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images, ISPRS J. Photogramm. Remote Sens. 116 (2016) 24–41.
[26] M. Roy, S. Ghosh, A. Ghosh, A novel approach for change detection of remotely sensed images using semi-supervised multiple classifier system, Inf. Sci. 269 (2014) 35–47.
[27] L. Xu, H. Zhang, M. Zhao, D. Chu, Y. Li, Integrating spectral and spatial features for hyperspectral image classification using low-rank representation, in: Proceedings of IEEE International Conference on Industrial Technology (ICIT), IEEE, 2017, pp. 1024–1029.
[28] W.A. Malila, Change Vector Analysis: An Approach for Detecting Forest Changes with Landsat, LARS Symposia, 1980, p. 385.

Chinmayee Pati received the B.Tech. degree in Computer Science & Engineering from College of Engineering Bhubaneswar under Bijupatnaik University of Technology in 2010 and the M.E. degree in Computer Science & Engineering from PG Department of Computer Science and Applications, Utkal University in 2013. Currently she is perusing Ph.D. degree in the department of Computer Science and Applications, Utkal University. Her research interest includes Image Processing, Machine Learning and Soft Computing.

Ashok Kumar Panda received the MCA degree in 1999 and M.Tech. (Computer Science) degree in 2006 both from Utkal University. In 2000, he joined in the MCA department of Institute of Science & Information Technology, Bhubaneswar where he currently is a Sr. Assistant Professor. He is also a research scholar of PG department of Computer Science and Applications, Utkal University. His research interest includes Image Processing, Machine Learning and Real Time Database Systems.

Dr. Ajaya Kumar Tripathy is working as an Associate Professor in department of Computer Science and Engineering, Silicon Institute of Technology, Bhubaneswar, India. Dr. Tripathy holds a Professional M.Tech. Degree in e-Government from Trento University, Italy in 2008. In 2007 he completed the M.Tech. Degree in Computer Science from Utkal University, India. From 2008 to 2010 he was a PhD student in the SOA Research Unit at Fondazione Bruno Kessler (FBK) in Trento, Italy. Dr. Tripathy also worked as a research assistant (internee) in Indian Statistical Institute, India and Createnet Research Center, Trento, Italy. His research interest include pattern recognition, machine learning, data mining and soft computing.

Dr. Sateesh Kumar Pradhan obtained his Ph.D. Degree in Computer Science (CS) from Berhampur University in 1999. He was Head, PG Dept. of CS, Utkal University in 2001-2003. He was the Organizing Chair of the International Conference on Information Technology-2005. He served as a senior faculty in the Computer Engineer ing Dept. of King Khalid University, Saudi Arab from 2006 to 2011. Currently he is a Professor in Graduate Department of Computer Science, Utkal University. His research interest includes Neural Computing, Computer Architecture, Ad hoc Networks, Computer Forensic.

Dr. Srikanta Patnaik is a Professor in the Department of Computer Science and Engineering, Faculty of Engineering and Technology, SOA University, Bhubaneswar, India. He has received his Ph. D. (Engineering) on Computational Intelligence from Jadavpur University, India in 1999 and supervised 15 Ph. D. Theses and more than 50 Master theses in the area of Computational Intelligence, Soft Computing Applications and Re-Engineering. Dr. Patnaik has published around 100 research papers in international journals and conference proceedings. He is author of 2 text books and 15 edited volumes and few invited book chapters, published by leading international publisher like Springer-Verlag, Kluwer Academic, etc.. Dr. Patnaik was the Principal Investigator of AICTE sponsored TAPTEC project "Building Cognition for Intelligent Robot" and UGC sponsored Major Research Project "Machine Learning and Perception using Cognition Methods". He is the Editors-in-Chief of International Journal of Information and CommunicationTechnology and International Journal of Computational Vision and Robotics published from Inderscience PublishingHouse, England and also Editors-in-chief of Book Series on "Modeling and Optimization in Science and Technology" published from Springer, Germany. He is also Series Editor of the book series of "Automation, Control and Robotics" published from River Publication, Netherland.