



# Multi-scale Rols selection for classifying multi-spectral images

Ayan Seal<sup>1</sup> · Angel Garcia-Pedrero<sup>2</sup> · Debotosh Bhattacharjee<sup>3</sup> · Mita Nasipuri<sup>3</sup> · Mario Lillo-Saavedra<sup>4</sup> · Ernestina Menasalvas<sup>2</sup> · Consuleo Gonzalo-Martin<sup>2</sup>

Received: 2 October 2018 / Revised: 7 September 2019 / Accepted: 30 September 2019 /  
Published online: 11 October 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

The applications of object-based image analysis (OBIA) in remote sensing studies have received a considerable amount of attention over the recent decade due to dramatically increasing of the spatial resolution of satellite imaging sensors for earth observation. In this study, an unsupervised methodology based on OBIA paradigm for the estimation of multi-scale training sets for land cover classification is proposed. The proposed method consists of selection of valid region of interests in an unsupervised way and its characterization using some attributes in order to form meaningful and reliable training sets for supervised classification of different land covers of a satellite image. Multi-scale image segmentation is a prerequisite step for estimation of multi-scale training sets. However, scale selection remains a challenge in multi-scale segmentation. In this work, we propose a method to determine the appropriate segmentation scale for each land cover with the help of prior knowledge in the form of in-situ data. The proposed method is further discussed and validated through multi-scale segmentation using quick shift and random forest algorithms on two multi-spectral images captured using Worldview-2 sensor. Experimental results indicate that the proposed method qualitatively and quantitatively outperforms three state-of-the-art methods.

**Keywords** Object based image analysis · Multi-scale RoIs · Quick shift · Random forest

## 1 Introduction

Identification of different land covers from remotely sensed satellite images is an essential task for earth observation (EO). It has many applications in geology and mineral exploration,

---

✉ Ayan Seal  
ayan.seal@gmail.com

<sup>1</sup> PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur 482005, India

<sup>2</sup> Center for Biomedical Technology, Universidad Politécnica de Madrid, Campus de Montegancedo, 28233 Pozuelo de Alarcón, Spain

<sup>3</sup> Jadavpur University, Kolkata, India

<sup>4</sup> Universidad de Concepción, Concepcion, Chile

hazard assessment, oceanography, agriculture and forestry, land degradation, environmental monitoring, and crop monitoring (Blaschke et al. 2008). Several works have been carried out so far to explore useful information related to EO based on the high spectral and/or spatial resolution of remote sensed images in the last decade. Most of these works are based on the classification of pixels by considering the spectral similarities with predefined land covers in early days because this approach has some strong advantages. However, this approach has several shortcomings (Fisher 1997; Carleer et al. 2005; Aplin and Smith 2008; Gao and Mas 2008) as well. In contrast to traditional pixel-based analysis, object-based image analysis (OBIA) considers “regions” or “objects” as basic building blocks for many tasks such as temporal change detection, target recognition, and land-cover classification to produce valuable geographic information. In this study, objects and regions are used synonymously. This approach provides numerous advantages (Blaschke et al. 2008; Blaschke 2010; Blaschke and Strobl 2001; Blaschke and Lang 2006; Blaschke et al. 2010), including reduced spectral variability and more spatial and contextual information such as shape and topological relationships. The main idea behind this approach is multi-scale image segmentation in which an image is partitioned into hierarchical homogeneous objects or regions by a top-down approach (Kurtz et al. 2012; Dey et al. 2010) or a bottom-up approach (Benz et al. 2004). OBIA mainly concentrates on different characteristics or attributes or features extracted from a region for its identification. Different features are region’s size, shape, texture, and colour. It also considers the contextual information of each region by evaluating the spatial relationships among the image regions at different scales (Blaschke 2010). The feature set is suitable for analyzing land covers using multi-scale image analysis, which matches the different scales with the sizes of the regions (Baatz and Shape 2000; Dungan et al. 2002; Hay and Marceau 2004). Many image segmentation algorithms used in remote sensing application are summarized and reviewed (Blaschke 2010; Blaschke et al. 2004; Dey et al. 2010; Mohan and Ladha 2009; UcaAvci et al. 2011; Gholoobi et al. 2010; Whiteside and Ahmad 2005; Garcia-Pedrero et al. 2015). In top-down approaches, an image is divided into regions based on some heterogeneity criteria whereas in bottom-up approach, objects are formed by merging pixels using some homogeneity criteria. The land cover classification accuracy does not depend on which segmentation method is used, but on the segmentation result and its scales. Currently, very few methods are there in literature for automatic selection of optimal scale for image segmentation (Kim et al. 2008). In most of the cases, a scale selection mechanism depends on subjective trial-and-error basis (Meinel and Neubert 2004; Zhang et al. 2008; Duro et al. 2012; Huang and Zhang 2008). Therefore, suitable scale selection is a challenging problem for multi-scale image segmentation (Ming et al. 2015; Drăguț et al. 2010.) In Kim et al. (2008), authors used local variance and spatial autocorrelation to find out the value of scale for forest classification from multi-spectral IKONOS images. Local variance was used further to develop a tool for estimating range of suitable scales across a range of different image types and landscapes (Drăguț et al. 2010). In Woodcock and Strahler (1987), authors used local variance graphs for defining scale of ground regions because local variance graphs can reveal the spatial structure of images. Intra-segment homogeneity using weighted variance of the near-infrared band (Espindola et al. 2006) and inter-segment heterogeneity using spatial autocorrelation (Johnson and Xie 2011) were exploited by two different groups to determine scale for image segmentation. In Karl and Maurer (2010), authors used variogram-based spatial dependency to estimate suitable scales for generating land-management information. The local-variance graph was considered to estimate spatial scale for mean-shift segmentation algorithm (Ming et al. 2015; Ming and Zhang 2016.) In Drăguț et al. (2010, 2011), Trias-Sanz et al. (2008), Zhou et al. (2017), authors developed an adaptive approach for scale selection in multi-scale segmentation of multi-spectral satellite images namely GF-1 and ZY-3 based on

inherent features of images and prior knowledge of thematic maps. In Gonzalo-Martín et al. (2013), authors obtained scales by post-evaluation of the multi-scale segmentation results using quick shift (QS) segmentation algorithm (Vedaldi and Soatto 2008) and decision tree (Tan 2018).

It is clear from the above discussion that the objective of most of these methods is to select optimal scale by analyzing either the statistical variance of some features of image before segmentation or multi-scale image segmentation results (Ming et al. 2015, 2012). Moreover, these methods derive the necessary features and measurements used in scale selection from the images. No prior knowledge is exploited except in Zhou et al. (2017), Witharana and Civco (2014), Drăguț et al. (2014).

The objective of this work is to propose an unsupervised methodology for the estimation of multi-scale training sets. The proposed method consists of selection of valid region of interests (ROIs) in an unsupervised way and followed by their characterization using some attributes in order to form meaningful and reliable training sets for supervised classification of different land covers of a satellite image. However, multi-scale image segmentation is a prerequisite step of estimation of multi-scale training sets. The prior knowledge in the form of in-situ data helps to determine the appropriate segmentation scale for each type of land cover. The proposed ROI selection method is further validated, both qualitatively and quantitatively through multi-scale segmentation using QS and random forest (RF) on two multi-spectral images. The results obtained by the proposed method are also compared with the results achieved using three state-of-the-art methods.

The rest of the article is as follows: In section 2, QS and RF are briefly reviewed. The proposed framework is presented in Sect. 3. Experimental results are reported in Sect. 4. Section 5 concludes the work.

## 2 Background

This work is carried out using existing QS and RF algorithms. Brief review of each of the algorithm is described as follows:

### 2.1 Quick shift algorithm

Segmentation refers to the process of partitioning an image into perceptually meaningful homogeneous and connected regions. A superpixel based algorithm can be used to replace the rigid structure of the pixel grid (Achanta et al. 2012). A superpixel representation significantly decreases the number of image primitives compared to the pixel representation. It also offers the spatial support for computing region based features, which are more meaningful and easier to analyze. There are several mode seeking image segmentation algorithms namely, Mean shift (MS) (Fukunaga and Hostler 1975), Medoid shift (MeS) (Sheikh et al. 2007) and QS (Vedaldi and Soatto 2008) in literature, which produce superpixels. But, MS and MeS algorithms do not have ability to control the size, number, and compactness of the regions. On the other hand, QS produces superpixels that are fixed in approximate size or number (Vedaldi and Soatto 2008). MeS is faster than MS if the underlying distance is Euclidean, but it fails to cluster data points belonging to the same mode, resulting in over-fragmentation. QS operates in non-Euclidean spaces like MeS and trades-offs between under-fragmentation and over fragmentation. It has the ability to identify clusters of pixels in the joint spatial and color dimensions even if an image has more than one channel or multiple channels. Segments are

considered as local elements (superpixels) and can be used as a basis for further processing. It takes lesser time for the segmentation process, since it does not consider each pixel point in the feature space. Generating superpixels using QS are controlled by three parameters namely, ratio( $r$ ), kernel size( $k$ ) and distance/scale( $d$ ). The two words namely, scale and distance are same in this context. So, these two words are used interchangeably in this work. Ratio is a trade-off between spatial importance ( $i, j$ ) and spectral importance ( $I^1, I^2, \dots, I^n$ ), where  $I^1$  is a panchromatic image and  $I^2$  to  $I^n$  are multi-spectral images. It balances multi-spectral band proximity and image-space proximity and has a value in the range [0, 1]. Small ratio corresponds to more importance to the spatial component than multi-spectral band. On the other hand, higher values give more weight to multi-spectral band. Scale concentrates on those points, which are similar regardless of their density. Here, the density represents standard deviation ( $\sigma$ ), which is estimated by the Parzen window density estimator. Higher density value means fewer clusters. For each pixel  $(x, y)$ , QS considers  $\{x, y, I^n(x, y)\}$  as a sample from a  $n+2$  dimensional vector space. Then, it calculates the Parzen density estimate with a Gaussian kernel of  $\sigma$  by Eq. 1.

$$E(x, y) = P\{x, y, I^n(x, y)\} = \sum_{x', y'} \frac{1}{(2\pi\sigma)^{n+2}} \exp\left(\left(-\frac{1}{2\sigma^2}\right) \begin{bmatrix} x - x' \\ y - y' \\ I^n(x, y) - I^n(x', y') \end{bmatrix} \right), \tag{1}$$

where  $x'$  and  $y'$  are the x-coordinate and y-coordinate of a pixel. In this window method, a tree is formed based on the estimated density of each point in the feature space, where the root having the highest density. A tree can be represented formally by Eq. 2.

$$(x', y') > P(x, y) \text{ iff } P(x', y', I^n(x', y')) > P(x, y, I^n(x, y)) \tag{2}$$

Scale is the maximum distance between nodes in the QS tree, which is used to cut links in the tree to form the segmentation. Each pixel  $(x, y)$  is connected to the closest higher density pixel parent  $(x, y)$  that achieves the minimum scale by equation 3.

$$dist(i, j) = \min_{(x', y') > P(x, y)} (x - x')^2 + (y - y')^2 + \sum_{i=1}^n \|I^n(x, y) - I^n(x', y')\|_2^2 \tag{3}$$

The algorithm calculates a forest of pixels whose branches are labeled with a scale value. This specifies a hierarchical segmentation of the image, with segments corresponding to subtrees. Useful superpixels can be identified by cutting the branches whose scale label is above a given threshold ‘ $d$ ’.

### 2.2 Random forest (RF)

In recent years, a great interest has grown on the RF for analysing data (Seal et al. 2016). It is an ensemble learning method. It generates many decision tree based classifiers and takes decision based on the aggregated results. Two popular ensemble learning methods are the boosting (Shapire et al. 1997) and bagging (Breiman 1996) of classification trees. In the boosting method, successive trees depend on earlier trees, and an extra weight is provided to the points which are incorrectly predicted by the earlier predictors. A weighted vote is then considered for prediction. On the other hand, successive trees do not depend on earlier

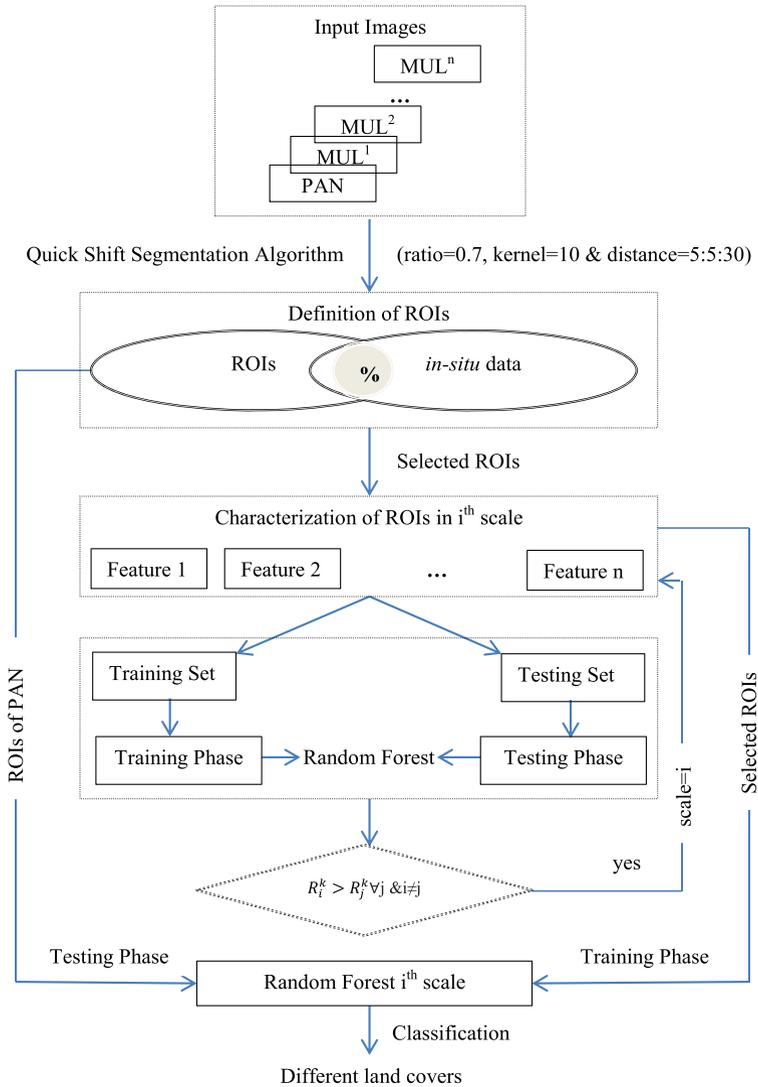
trees in bagging method. The RF employs in bagging method (Breiman 2000, 2001, 2004, 2017), where each tree is constructed independently using different subsets of data sets with replacement. The data is selected randomly and uniformly. It may happen that some data set may be picked twice and other data set may not be picked. The idea of drawing at random data set and then using different data sets for the construction of classification trees is known as bootstrapping. Finally, the predictions of all trees are aggregated using the majority voting for classification.

### 3 Proposed methodology

The block diagram of the proposed system is shown in Fig. 1, where selection of scale is carried out manually and automatically. The first step of the proposed system is multi-scale image segmentation, which takes panchromatic (PAN) and ‘n’ number of spectral bands ( $MUL^n$ ) of a satellite image as input. The segmentation algorithm splits the input image into objects. Then each segmented region is checked whether it overlaps with any of the labeled land cover regions in the in-situ data. If the overlap is more than or equal to 90% of region’s own area then it is treated as a valid region or region of interest (RoI) and it is assigned the label of that land cover region. These RoIs are used further for the estimation of multi-scale training sets in order to classify different land covers in an image. In this work, QS is adopted as the segmentation algorithm. The size of each region depends on the scale parameter. In manual approach, as the value of the scale parameter is not known, QS is applied on same set of satellite images several times using various scale parameters, generating multi-scale segmentation results. Each of the segmented RoIs is then characterized by extracting the following features from it.

- (i) Mean gray values of the pixels constituting each RoI for each band of multi-spectral images and panchromatic image,
- (ii) Standard deviation of the gray levels of the pixels constituting each RoI for each band of multi-spectral images and panchromatic image,
- (iii) Entropy of the gray levels of the pixels constituting each RoI for each band of multi-spectral images and panchromatic image,
- (iv) Some derived features from multi-spectral bands such as
  - (a) Normalized difference vegetation index (NDVI) (Geerken et al. 2005),
  - (b) Normalized difference water index (NDWI) (Gao 1996), and
  - (c) Non-homogeneous feature difference (NHFD) (Bhattarai et al. 2011).

It is observed that the above mentioned features are sufficient to classify land covers in satellite images. Thus, features like textural, shape feature, context/neighbourhood, arrangement of objects etc are not used in this study. These features are calculated for each segmented RoI in each scale independently. It means that the number of feature sets is equal to the number of different scales that are considered. These labeled RoIs help to train RF for classifying different land covers in a satellite image. When the RoIs and their feature extraction is over at each scale then “the best scale” named as ‘scale=i’ in Fig. 1 for each type of land cover has to be found, which help to build a multi-scale RoIs set for the training process of RF to recognize different land covers in satellite images that are not involved in the training process. The manual scale determination process starts with randomly dividing the selected labeled RoIs into two sets namely, training and test set for each scale separately. The training set is fed into fifteen RF classifiers for each scale separately. After the training is finished,



**Fig. 1** Schematic block diagram for land covers classification when multi-scale ROIs are chosen manually for training

corresponding test set is fed to the trained model for identifying the class label of each test sample. The scale is considered to be the best, which corresponds to the higher accuracy performance in a classification process. Lastly, the “best scale” named as ‘scale=i’ of a particular land cover, ‘k’, is chosen depending on the highest accuracy ( $R_i^k > R_j^k \forall j \& i \neq j$ ) of that particular land cover among all scales. Here,  $R_i^k$  is the accuracy rate of the *i*th scale for land cover ‘k’ and ‘j’ is also used to represent scale. Based on the “best scale”, ‘scale=i’ for a particular land cover ‘k’, final training set forms, which is used further for the classification of land cover of panchromatic image.

### 3.1 Determination of scale

The manual scale selection method has some limitations. Firstly, the QS algorithm depends on scale, which needs to be provided as input by a user. Thus, it is not an automatic approach. Moreover, the identification of such scale for each land cover is not always an easy task. Secondly, sometimes it may happen that at a particular scale, the RoIs generated may not have the requisite (90%) overlap with the smallest labeled polygon in the in-situ data. This happens when the value of scale is larger than the radius of land cover. To overcome these limitations, a method is proposed in this work, which finds scale for each land cover automatically. The prior knowledge in the form of in-situ data helps to know the appropriate scale for each land cover. *The in-situ data consists of some labeled polygons representing different land covers.* Moreover, in-situ data also consists of more than one polygon for depicting same land cover. Since the areas of the different polygons associated to a same land cover are not equal, the proposed algorithm for automatic scale selection is based on the smallest polygon of a particular land cover. In this way, multiple scales are chosen for RoIs representing different land covers which are used further for training process of RF. The steps of the proposed method are as follows.

All the polygons representing different land covers in-situ data are convex in nature. Thus, there is no chance having its centroid out of the polygonal region. Here, a land cover or polygon is treated as a class.

1. The total number of polygons representing each class is to be identified from the given situ data.
2. Determination of scale for each class.
  - 2.1) There may be more than one polygon of a particular class. So, find the smallest polygon in terms of its area. Before finding the smallest polygon, the image needs to be converted into its binary form, where ‘1’ corresponds to existence of a class and it is represented by white color and ‘0’ means the background, which is specified by the black color in the same figure. The area is the total number of white pixels in a polygon. So, find the smallest polygon based on the total number of white pixels from the many polygons of a particular land cover.
  - 2.2) Then, find the centroid of this polygon using Eqs. 4 and 5 (Seal et al. 2015).

$$X = \frac{\sum f(x, y)x}{\sum f(x, y)}, \quad (4)$$

$$Y = \frac{\sum f(x, y)y}{\sum f(x, y)}, \quad (5)$$

where  $X$  and  $Y$  are the co-ordinates of the centroid of the smallest polygon and  $x, y$  are the co-ordinates of the binary image or a class and  $f(x, y)$  is the intensity value that is either ‘1’ or ‘0’.

2.3) Finally, Euclidean distances between the centroid and all boundary points of the smallest polygon have to be estimated. Actual distance would be the shortest distance among all and divide it by constant ‘4’, which is denoted as ‘ $d^*$ ’.

\* Note: Many experiments have been performed in order to find the proper constant term in step 2.3 for all the classes and finally the value ‘4’ is chosen. We believe that this constant factor depends on size of the land covers. Moreover, it depends on size of the image.

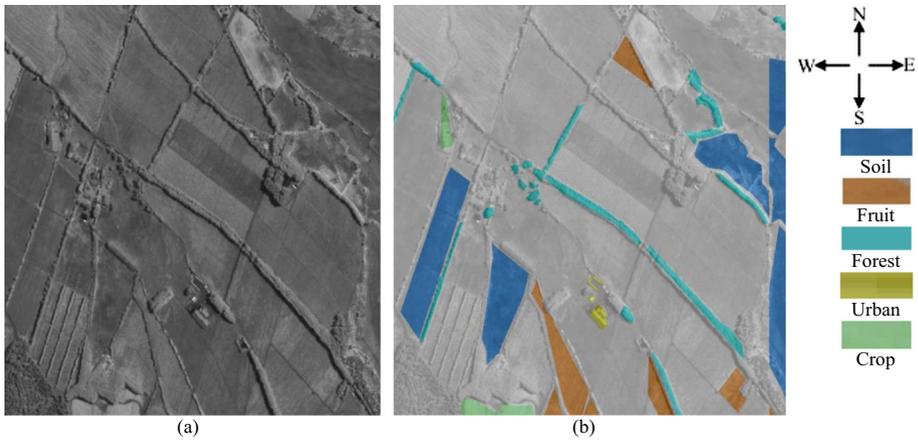
## 4 Experimental results and discussion

In this work, the proposed method is evaluated on two different sets of multi-spectral and panchromatic images along with their in-situ data. It consists of information related to ground truth, which helps to assign the class label of RoIs obtained using QS for training. Both the sets are captured using Worldview-2 sensor at 1:25,000 scale. Satellite image data and spectral measurements were correlated and satellite were classified to determine the location within the studied region. The first set of images consists of one high resolution 0.46m panchromatic band and 8 multi-spectral bands of 1.8m resolution. The multi-spectral band consists of red, blue, green, near infrared, red-edge, coastal, yellow and near infrared 2. The first four are the standard bands while the others are new bands (Globe 2011). The size of each image is  $2048 \times 2048$  pixels. These images were acquired on 11<sup>th</sup> September, 2011. The images covered an area of 10.485 ha and the coordinates of upper left corner are  $S32^{\circ}51'7.91''$  and  $W70^{\circ}39'5.10''$  respectively. The area corresponds to a rural zone located at Valparaiso region in Comuna de Los Andes, Chile. Seven land covers were found in the studied area in dataset\_1. These land covers are generic agricultural land, water bodies, and four different types of crops (nectarine, grapevine, alfalfa and maize) in different phenological stages and buildings and urban construction. The nectarine crops area could be separated further in two different crop areas: nectarin\_1 and nectarin\_2. The panchromatic image of dataset\_1 is shown in Fig. 2a.

The second set of images was captured on 19<sup>th</sup> January, 2012 using the same sensor. The images are covered in an area of 157 ha of croplands at Coihueco district, in Nuble province, Biobio region, Chile ( $S36^{\circ}37'15.7''$  and  $W71^{\circ}53'57.7''$ ). The area is a good representation of diverse vegetation, forests, rural constructions, and crops. The satellite images consist of one panchromatic image of 0.59 m resolution and 4 multi-spectral of resolution 2.36 m. Four spectral bands are as follows: blue band (450–510 nm), green band (510–580 nm), red band (630–690 nm), and near infrared band (NIR, 770–895 nm). The size of each image is  $2006 \times 2172$  pixels. Five land covers were found in dataset\_2. These land covers are forest, soil, crop, fruit, and urban construction. The panchromatic image of dataset\_2 is shown in Fig. 3a. The labeled polygons included in the vector file are overlaid on Figs. 2a and 3a,



**Fig. 2** **a** Panchromatic image of dataset\_1 and **b** corresponding in-situ data showing selected regions of different types of land covers with color labels

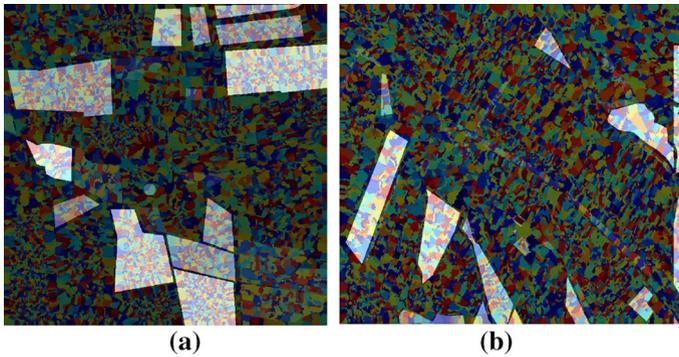


**Fig. 3** **a** Panchromatic image of dataset\_2 and **b** corresponding in-situ data showing selected regions of different types of land covers with color labels

shown in Figs. 2b and 3b respectively. Beside these in-situ data, pixel-wise ground truth of land cover types are also provided with both the datasets.

The QS may be used for any feature space, but for this work, the segmentation process using QS is done based on the raw intensity values of panchromatic band and multi-spectral bands of satellite images. Two set of experiments are carried out in this work. As discussed before, the QS algorithm depends on three parameters namely, ‘r’, ‘k’ and ‘d’. The value of the ‘r’ should be close to ‘1’ for this work for assigning more weight to multi-spectral space. Thus, three loops are used for finalizing the values of ‘r’, ‘k’ and ‘d’. Outer loop refers to ‘r’ and two inner loops are used for other two parameters. Initially, the values of ‘r’, ‘k’ and ‘d’ are ‘1’, ‘1’ and ‘1’ respectively. Then the value of ‘r’ is decremented by 0.1 and the values of the other two parameters are incremented by ‘1’ in each successive iteration. After analyzing the outputs of the QS, the value of ‘k’ is set to ‘10’ because  $k = 10$  provides a mean segment size for the lower value of ‘d’ adequate for the scenes under study. But, the identification of scale parameter is not an easy task. In the first set of experiments, the scales are predefined i.e. in other words, the values of scales are chosen manually, which are 5, 10, 15, 20, 25, and 30. So, the segmentation process is carried out six times depending on these six values of scale, where scale is measured in numbers of pixels.

Figure 4 demonstrates the output of the QS algorithm for  $r = 0.7$ ,  $k = 10$ , and  $d = 5$  which are overlaid on different land covers obtained from in-situ data for both set of images. The outputs of the QS, i.e. RoIs at each scale are marked by randomly assigned colors. After that some of the segmented regions are selected by a simple rule, which is also described in Sect. 2.1. Then the selected RoIs are characterized using the designed features and are fed into RF for training and testing. The selected RoIs are divided into two random sets, one containing 70% of the selected RoIs and other containing rest 30%. It is found experimentally that the accuracy is higher when the size of the training set and test set are 70% and 30% of selected RoIs respectively. However, accuracy increases marginally if 75% RoIs are used for training process. But, it takes more time than the previous case. That is why the rest of the experiments are based on 70% RoIs and 30% RoIs for training and testing processes respectively. Training set and their labels are used to train the fifteen RF classifiers. In general, the better results depend on the use of more number of trees. After a certain point, the improvement decreases



**Fig. 4** **a** RoIs obtained from QS with  $d = 5$  for the scene in Fig. 2 with overlaid labeled RoIs and **b** RoIs obtained from QS with  $d = 5$  for the scene in Fig. 3 with overlaid labeled RoIs

**Table 1** Recognition accuracies (%) for different classes of dataset\_1 at various scales

Scale	Alfalfa	Water	Urban	Maize	Nectarin_1	Nectarin_2	Grapevine	Soil
5	98.91	<b>99.63</b>	84.63	95.42	<b>98.39</b>	<b>81.02</b>	98.74	92.03
10	99.01	92	81.33	<b>95.78</b>	98.35	76.16	99.07	95.82
15	<b>99.26</b>	94.52	<b>86.23</b>	94.40	98.14	72.69	<b>99.19</b>	94.56
20	99.12	Null	Null	94.96	97.70	74.67	98.82	94.46
25	98.95	Null	Null	Null	97.69	79.31	98.92	<b>98.87</b>
30	99.10	Null	Null	Null	96.33	73.85	98.43	98.12

as the number of trees increases. In other words, the advantage in prediction from learning more trees will be lower than the cost in computation time for learning these additional trees. It is observed experimentally that the accuracy decreases as the number of trees increases after fifteen. After training process, test set is used for identifying the class label of each test sample. The accuracies (Karlekar et al. 2019; Sharma and Seal 2019) (%) of two datasets are shown in Tables 1 and 2 respectively. Each row of the Tables 1 and 2, corresponds to the accuracies based on the RoIs from the segmented image using different scales. The null value represents that there are no valid RoIs for the particular land cover for that particular scale. It happens when the value of scale is larger than the radius of land cover. The best accuracy of each land cover is marked in bold font in Tables 1 and 2 and corresponding scale is noted. After selecting the best scale for a particular type of land cover, all the segmented regions at that scale are fed to the selected classifier and the accuracy performance of the classifier is noted by comparing with the pixel-wise labeling of the land cover type provided as ground truth with the dataset.

Two confusion matrices are generated based on this actual outcome of the classifiers at the best scales for dataset\_1 and dataset\_2. These two confusion matrices are shown in Tables 3 and 4 for different land covers of dataset\_1 and dataset\_2 respectively when multi-scale RoIs are selected manually. Each column of the confusion matrix represents the instances in an actual class while each row represents the instances in a predicted class (Seal et al. 2016).

Tables 5 and 6 report true positive (TP), false negative (FP), true negative (TN), false negative (FN), accuracy (Powers 2011), sensitivity, specificity and average accuracy for different classes of dataset\_1 and dataset\_2 respectively when multi-scale RoIs are chosen

**Table 2** Recognition accuracies (%) for different classes of dataset\_2 at different scales

Scale	Forest	Urban	Minor crop	Fruit	Soil
5	76.05	98.50	78.64	79.21	96.81
10	75.11	98.95	80.73	<b>81.35</b>	97.65
15	<b>89.54</b>	<b>99.83</b>	78.30	71.60	98.03
20	88.26	98.82	80.80	66.95	98.61
25	84.20	99.59	<b>87.84</b>	65.36	<b>99.19</b>
30	81.20	Null	Null	68.11	98.75

**Table 3** Confusion matrix (%) generated from the accuracy performance at the best scale selected manually for each of the land covers for dataset\_1

Actual class	Predicted class							
	Alfalfa	Water	Urban	Maize	Nectarin_1	Nectarin_2	Grapevine	Soil
Alfalfa	99.26	0.03	0.79	0.45	0.09	2.36	0.22	0.13
Water	0.13	99.63	1.20	0.77	0.23	4.62	0.17	0.09
Urban	0.04	0.05	86.23	0.56	0.11	2.00	0.03	0.26
Maize	0.17	0.09	2.27	95.78	0.07	3.99	0.05	0.12
Nectarin_1	0.08	0.01	3.03	0.46	98.39	1.92	0.03	0.07
Nectarin_2	0.23	0.15	3.91	0.97	0.43	81.02	0.09	0.06
Grapevine	0.03	0.02	1.33	0.87	0.35	3.02	99.19	0.40
Soil	0.06	0.02	1.24	0.14	0.33	1.07	0.22	98.87

**Table 4** Confusion matrix (%) generated from the accuracy performance at the best scale selected manually for each of the land covers for dataset\_2

Actual class	Predicted class				
	Forest	Urban	Minor crop	Fruit	Soil
Forest	89.54	0.03	2.16	4.61	0.16
Urban	5.09	99.83	3.38	5.02	0.09
Minor crop	0.39	0.04	87.84	5.89	0.35
Fruit	2.33	0.08	4.19	81.35	0.21
Soil	2.65	0.02	2.43	3.13	99.19

manually. The higher sensitive values for different classes of both the datasets rarely overlook a positive and higher specific value rarely consider a positive for anything that is not the target of testing.

Till now, the class labels of some of the RoIs that belong to the testing set are found and matched with the predefined class labels. In order to get the class label of the image or all RoIs in an image, a training set is needed with high discriminating capability. This training set is formed on the basis of the best features that are obtained from the multi-spectral image based on the best scale for each land cover. After getting all the valid RoIs and its corresponding features for the best scale, these are put together to build a training set. Now, the main job is to find the known and unknown land covers from the multi-spectral image i.e. the image is used for testing purpose. The first set of experiments is biased towards scales estimation. It is already mentioned that the proper selection of scale for every land cover in an image is truly a difficult task.

**Table 5** Different classification measures (%) derived from Table 3 for different classes of dataset\_1 when multi-scale RoIs are chosen manually

	Alfalfa	Water	Urban	Maize	Nectarin_1	Nectarin_2	Grapevine	Soil
TP	99.26	99.63	86.23	95.78	98.39	81.02	99.19	98.87
TN	95.93	92.79	96.96	93.24	94.4	94.16	93.98	96.92
FP	0.74	0.37	13.77	4.22	1.61	18.98	0.81	1.13
FN	4.07	7.21	3.03	6.76	5.6	5.84	6.02	3.08
Accuracy	99.39	99.05	97.89	98.62	99.09	96.89	99.14	99.59
Sensitivity	96.06	93.25	96.60	93.40	94.61	93.27	94.27	96.97
Specificity	99.89	99.94	98.06	99.39	99.76	97.33	99.88	99.83
Average accuracy	<b>98.70</b>							

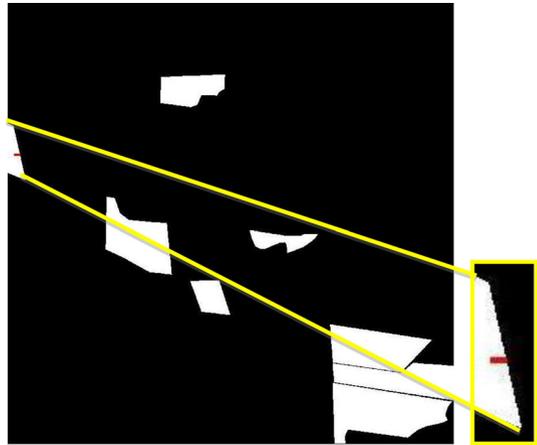
**Table 6** Different classification measures (%) derived from Table 4 for different classes of dataset\_2 when multi-scale RoIs are chosen manually

	Forest	Urban	Minor crop	Fruit	Soil
TP	89.54	99.83	87.84	81.35	99.19
TN	93.04	86.42	93.33	93.19	91.77
FP	10.46	0.17	12.16	18.65	0.81
FN	6.96	13.58	6.67	6.81	8.23
Accuracy	96.51	97.25	96.23	94.90	98.19
Sensitivity	92.78	88.02	92.94	92.27	99.33
Specificity	97.40	99.95	97.00	98.29	97.94
Average accuracy	<b>96.61</b>				

In the next set of experiments, the proposed algorithm is applied to automatically find the scale for each land cover present in a satellite image. The value of the scale will be higher for the segmentation with the increase in the area of the land cover or polygon in-situ data. The pictorial representation of the proposed algorithm for scale selection is shown in Fig. 5 for “Alfalfa” land cover of dataset\_1.

The calculated scales using proposed algorithm for dataset\_1 and dataset\_2 are shown in Tables 7 and 8 respectively. Next objective is to form a training set for building a trained RF, which will help to classify land covers of a satellite image, whose class labels are assumed to be unknown. Now, segmentation process is applied several times based on number of land covers present in a satellite image using automatic scales. For example, QS is applied 8 times for dataset\_1 whereas 5 times for dataset\_2 because 8 and 5 different land covers present in dataset\_1 and dataset\_2 respectively. Automatically, 8 different feature sets for dataset\_1 and 5 different feature sets for dataset\_2 are obtained based on 8 and 5 different scales. Similarly, 8 and 5 accuracies are obtained for each land cover using 8 and 5 different scales for dataset\_1 and dataset\_2 respectively. Out of these accuracies, only best accuracy for each land cover are reported in Tables 7 and 8. Two confusion matrices are shown in Tables 9 and 10 for different land covers of dataset\_1 and dataset\_2 respectively when multi-scale RoIs are selected automatically.

**Fig. 5** The smallest polygon of 'Alfalfa' land cover and its shortest scale



**Table 7** The automatic scale selection for different classes and the accuracy for the valid RoIs in each class of dataset\_1

Alfalfa	Water	Urban	Maize	Nectarin_1	Nectarin_2	Grapevine	Soil
<i>Scale accuracy (%)</i>							
7	3	9	16	13	7	30	24
98.21	97.48	98.15	98.12	98.15	98.24	98.01	98.21

**Table 8** The automatic scale selection for different classes and the accuracy for the valid RoIs in each class of dataset\_2

Forest	Urban	Minor Crop	Fruit	Soil
<i>Scale accuracy (%)</i>				
11	15	25	10	25
90.18	94.20	96.58	93.09	96.61

**Table 9** Confusion matrix (%) based on automatic multi-scale RoIs for different land covers of dataset\_1

Actual class	Predicted class							
	Alfalfa	Water	Urban	Maize	Nectarin_1	Nectarin_2	Grapevine	Soil
Alfalfa	98.21	0.59	0.47	0.17	0.32	0.08	0.17	0.09
Water	0.09	97.48	0.32	0.20	0.21	0.16	0.29	0.35
Urban	0.23	0.82	98.15	0.08	0.47	0.06	0.23	0.18
Maize	0.29	0.15	0.37	98.12	0.37	0.09	0.11	0.43
Nectarin_1	0.18	0.26	0.21	0.43	98.15	0.43	0.24	0.22
Nectarin_2	0.46	0.14	0.23	0.27	0.08	98.24	0.48	0.29
Grapevine	0.19	0.37	0.17	0.70	0.17	0.73	98.01	0.23
Soil	0.35	0.09	0.08	0.03	0.23	0.21	0.47	98.21

**Table 10** Confusion matrix (%) based on automatic multi-scale RoIs for different land covers of dataset\_2

Actual class	Predicted class				
	Forest	Urban	Minor crop	Fruit	Soil
Forest	90.18	1.19	0.87	1.93	0.37
Urban	2.17	94.20	0.59	1.23	1.27
Minor crop	3.35	2.33	96.58	1.69	0.83
Fruit	2.10	0.98	0.93	93.09	0.92
Soil	2.20	1.30	1.03	2.06	96.61

**Table 11** Different measures (%) derived from Table 9 for different land covers of dataset\_1

	Alfalfa	Water	Urban	Maize	Nectarin_1	Nectarin_2	Grapevine	Soil
TP	98.21	97.48	98.15	98.12	98.15	98.24	98.01	98.21
TN	99.34	98.38	97.93	98.19	98.03	98.05	97.44	98.54
FP	1.79	2.52	1.85	1.88	1.85	1.76	1.99	1.79
FN	0.66	1.62	2.07	1.81	1.97	1.95	2.56	1.46
Accuracy	99.69	99.48	99.51	99.53	99.52	99.53	99.43	99.59
Sensitivity	99.33	98.36	97.93	98.18	98.03	98.05	97.45	98.53
Specificity	99.74	99.64	99.73	99.73	99.73	99.72	99.71	99.74
Average accuracy	<b>99.53</b>							

**Table 12** Different measures (%) derived from Table 10 for different land covers of dataset\_2

	Forest	Urban	Minor crop	Fruit	Soil
TP	90.18	94.20	96.58	93.09	96.61
TN	95.64	94.74	91.80	95.07	93.41
FP	9.82	5.80	3.42	6.91	3.39
FN	4.36	5.26	8.20	4.93	6.59
Accuracy	97.16	97.78	97.67	97.63	98.04
Sensitivity	95.38	94.71	92.17	94.97	93.61
Specificity	97.57	98.55	99.13	98.28	99.14
Average accuracy	<b>97.65</b>				

Tables 11 and 12 present TP, FP, TN, FN, accuracy, sensitivity, specificity and average accuracy for different classes of dataset\_1 and dataset\_2 respectively when multi-scale RoIs are chosen automatically.

In order to compare the performance of the two RoIs selection methods, the accuracy provided for each of them for each land cover, and for the two studied data sets, are included in Figs. 6 and 7 respectively. Red color is used to represent the automatic method whereas dotted blue color depicts manual process. It is clear from both the Figs. 6 and 7 that the automatic selection of multi-scale RoIs results in higher accuracy than manual process for almost all the land covers.

The sensitivity and specificity are calculated for measuring the performance of the two RoIs selection methods. For the comparison of the performances of two RoIs selection methods, the sensitivity and specificity for each land cover of dataset\_1 are shown in Figs. 8 and 9,

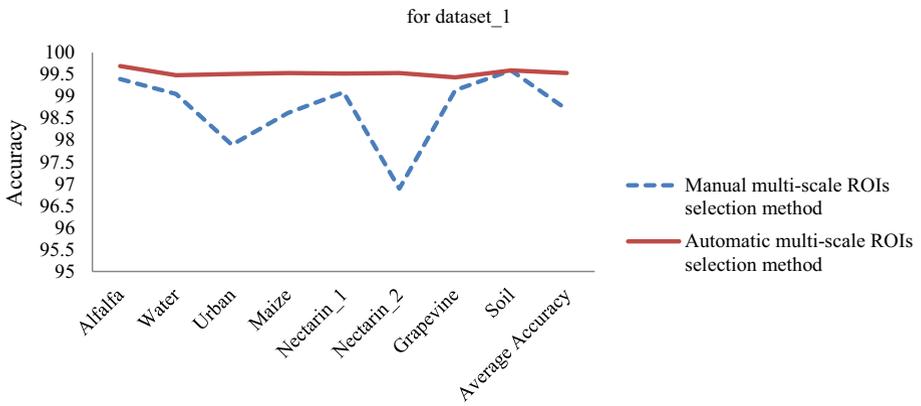


Fig. 6 Comparison of accuracies of manual and automatic multi-scale ROIs selection methods for dataset\_1

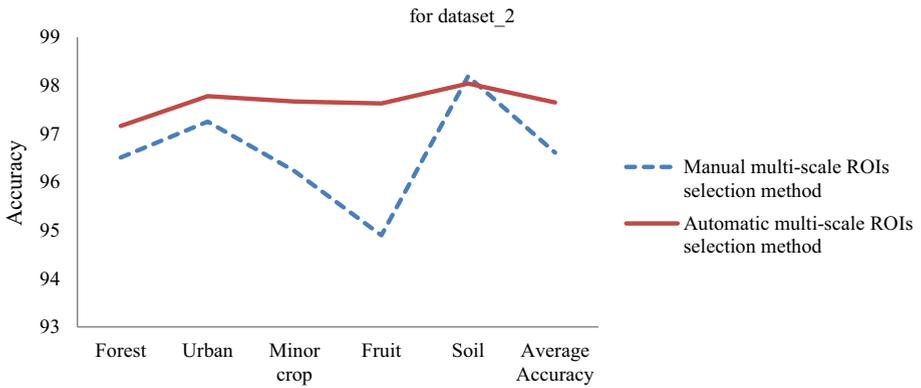


Fig. 7 Comparison of accuracies of manual and automatic multi-scale ROIs selection methods for dataset\_2

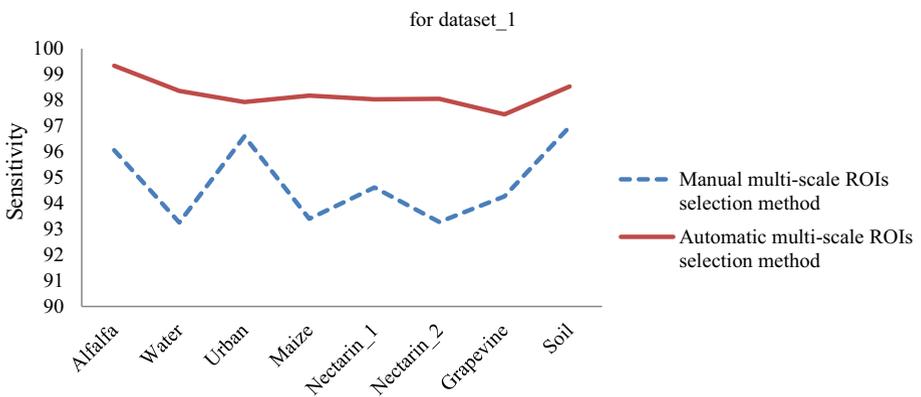


Fig. 8 Comparison of sensitivities of manual and automatic multi-scale ROIs selection methods for dataset\_1

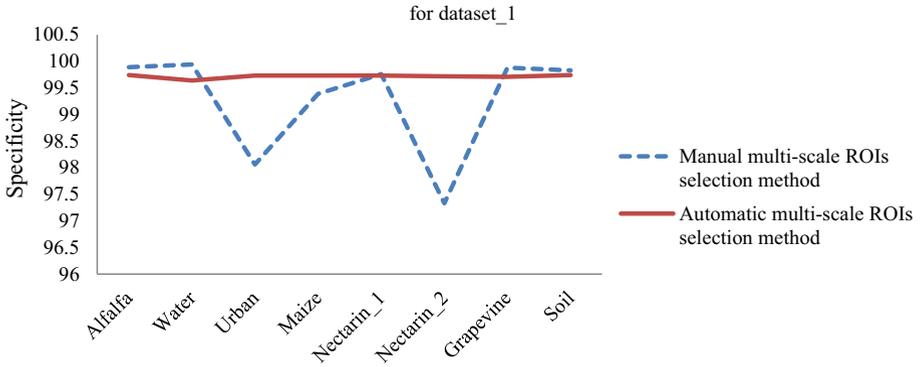


Fig. 9 Comparison of specificities of manual and automatic multi-scale ROIs selection methods for dataset\_1

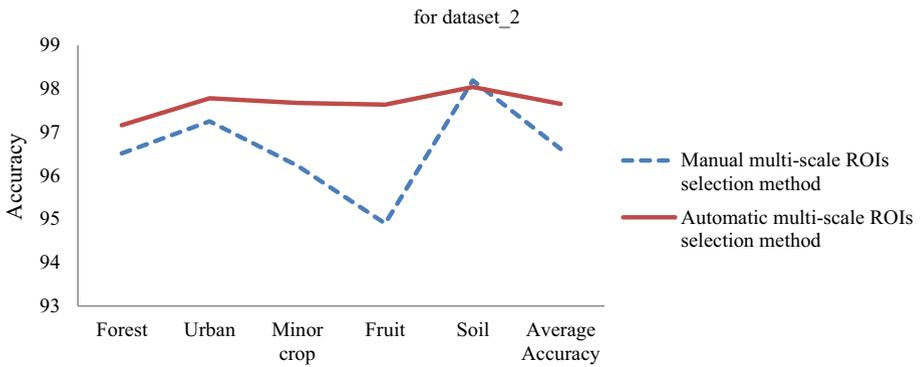


Fig. 10 Comparison of sensitivities of manual and automatic multi-scale ROIs selection methods for dataset\_2

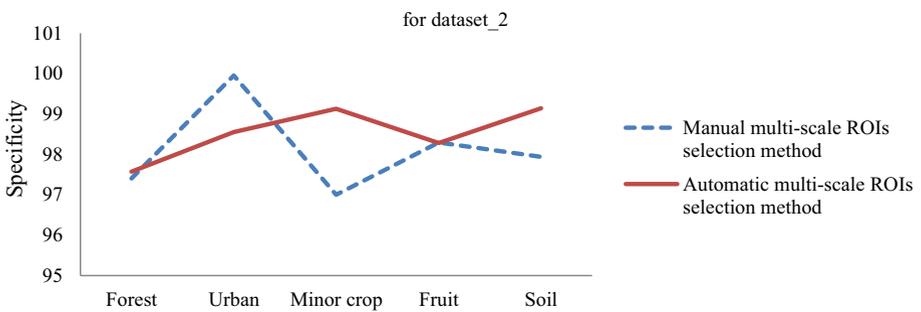
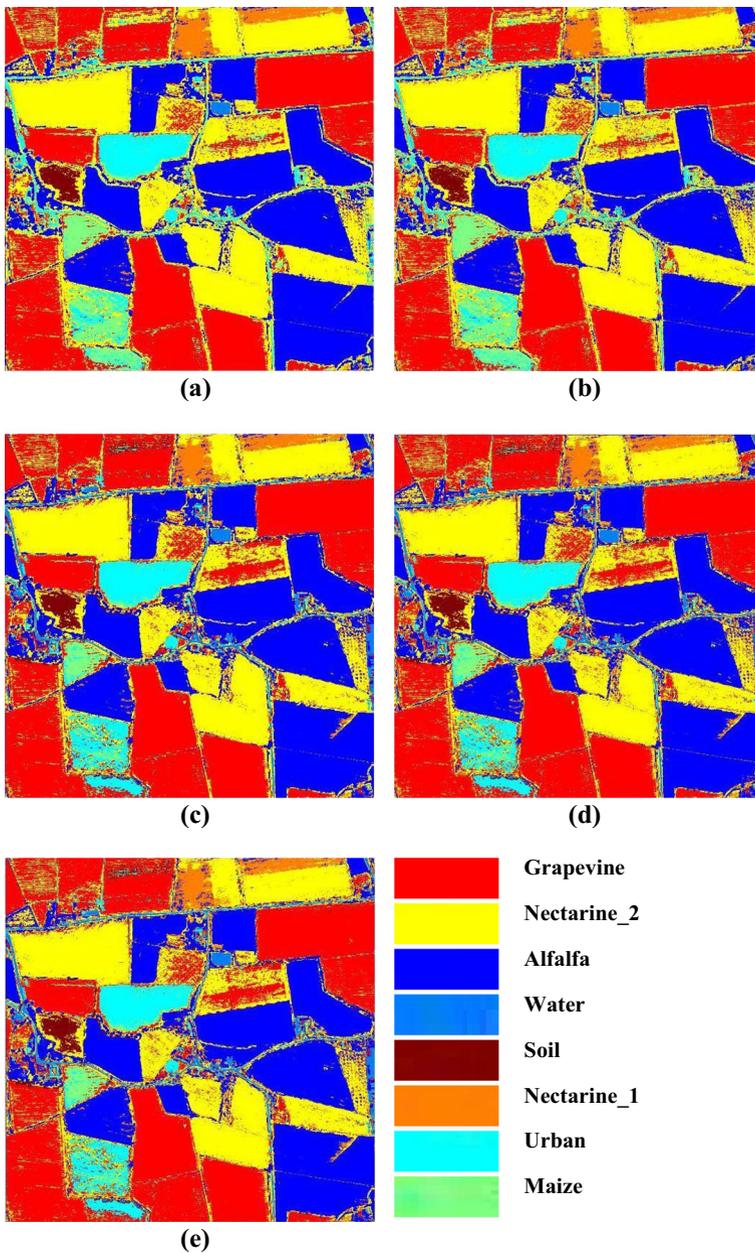
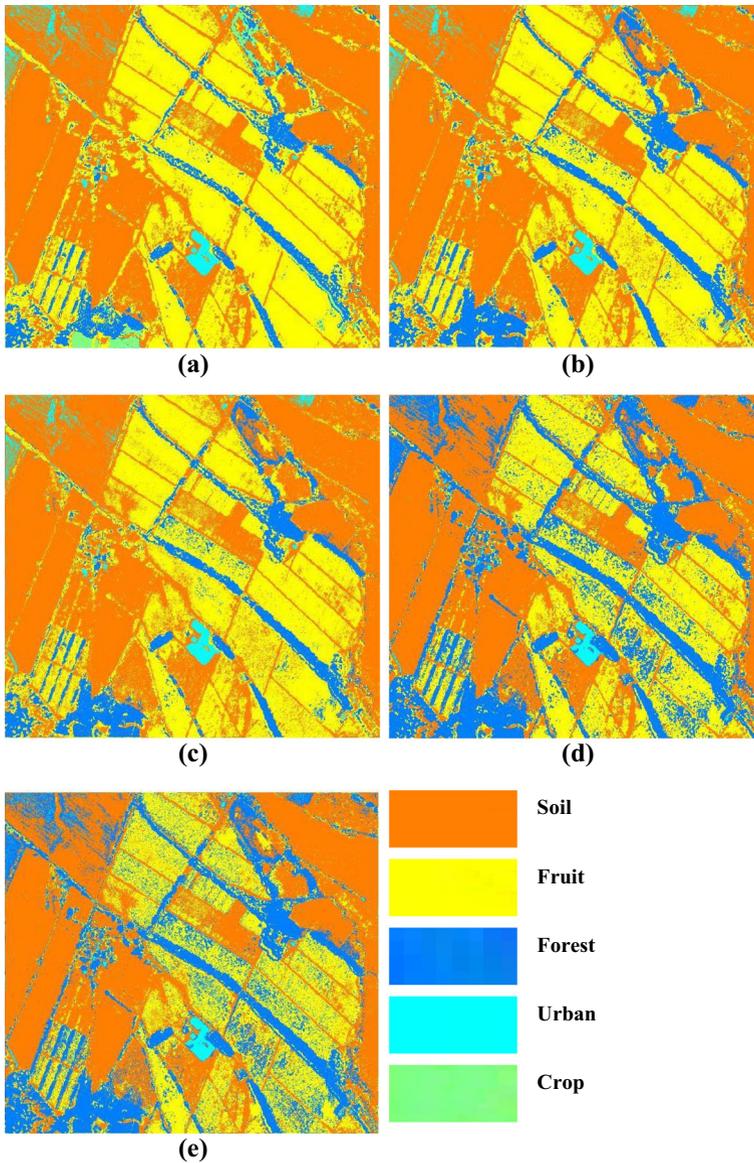


Fig. 11 Comparison of specificities of manual and automatic multi-scale ROIs selection methods for dataset\_2

whereas Figs. 10 and 11 depict the sensitivity and specificity for each land cover of dataset\_2. Red color is used to represent the automatic method whereas dotted blue color depicts manual process. It is clearly visible in Figs. 8, 9, 10 and 11 that automatic selection of multi-scale ROIs gives higher sensitivity and specificity than manual process for almost all the land covers.



**Fig. 12** Classification of different classes for dataset\_1 using **a** automatic scale selection method, **b** manual scale selection method, **c** M1 (Geiß et al. 2016), **d** M2 (Zhou et al. 2017) and **e** M3 (Wang et al. 2018)



**Fig. 13** Classification of different classes for dataset\_2 using **a** automatic scale selection method, **b** manual scale selection method, **c** M1 (Geiß et al. 2016), **d** M2 (Zhou et al. 2017) and **e** M3 (Wang et al. 2018)

The classification results of all the images namely, dataset\_1 and dataset\_2, using newly formed multi-scale RoIs are shown in Figs. 12a, b and 13a, b respectively. Some pseudo colors are used to represent different classes in both the datasets. Visually we can say that the classification results are good because more homogeneous areas are obtained with the automatic scale selection method. From the above obtained results, it is clear that the accuracies

**Table 13** The values of different metrics obtained by the two state-of-the-art methods and the proposed method

Method	Evaluation metric (%)	Dataset_1	Dataset_2
M1 (Geiß et al. 2016)	Average accuracy	82.24	81.76
	Average sensitivity	81.28	81.27
	Average specificity	82.17	82.30
M2 (Zhou et al. 2017)	Average accuracy	90.42	91.36
	Average sensitivity	91.26	91.78
	Average specificity	91.87	92.94
M3 (Wang et al. 2018)	Average accuracy	84.33	82.45
	Average sensitivity	83.33	82.68
	Average specificity	83.33	81.76
Proposed method	Average accuracy	99.53	97.65
	Average sensitivity	98.23	94.17
	Average specificity	99.72	98.53

are higher for each land cover when scales are chosen automatically. Moreover, it reduces human effort.

Now, the proposed approach is compared with three state-of-the-art approaches. In the first approach (M1), Geiß et al. (2016) used a bottom-up region-growing segmentation algorithm for partitioning remotely sensed image into objects. These objects were employed to combine several sequences secured with morphological operator based on central tendency. Then these objects were characterized by various shape features and fed into random forest classifier for land covers recognition. In this study, the same configuration is followed while implementing this approach on dataset\_1 and dataset\_2, where the achieved results are shown in Figs. 12c and 13c respectively. Further, the obtained average accuracies, average sensitivities and average specificities of the two datasets are reported in Table 13.

In second approach (M2), Zhou et al. (2017) merged features from each segmented region with prior knowledge from thematic maps in a top-down segmentation. First MS was used to find segmented regions from the whole image. Then three spectral features such as standard deviation, heterogeneous, and entropy, five textural features named as energy, contrast, entropy, inverse difference moment, and correlation, and two shape features including smoothness and shape compactness from each segmented region are considered. They computed complexity value for each region. A region would be further divided into multiple small regions if the complexity value of that region is large. The same procedure was followed to split into simplest regions iteratively. At last, the final segmentation results were obtained and were reported in Table 13. Figures 12d and 13d show the subjective evaluations on dataset\_1 and dataset\_2 respectively. In third approach (M3), Wang et al. (2018) considered a simple linear iterative clustering algorithm to find superpixels of image. Then these superpixels were partitioned by a dynamically constrained agglomerative clustering and partitioning (REDCAP) algorithm with the help of the initial number of segments. The local variance (LV) and the rate of LV change helped to determine suitable number of segments. Lastly, same superpixels were again repartitioned by the REDCAP algorithm according to the appropriate number of segments to acquire the segmentation result. The same work is adopted and is implemented on dataset\_1 and dataset\_2 and the obtained outcomes are displayed in Figs. 12e and 13e respectively. Furthermore, average accuracies, sensitivities and

specificities are stated in Table 13. It is clear from both the Figs. 12 and 13 and Table 13 that the proposed method i.e. automatic scale selection method outperforms three state-of-the-art methods qualitatively and quantitatively.

## 5 Conclusion

An automatic selection of multi-scale RoIs for multi-spectral satellite image classification is proposed in this work. The novelty of this work relies on the automatic selection of the scale parameter for the QS segmentation algorithm which is based on the labeled land covers of the in-situ data. Finally, the experiments are conducted using the RF method. Mainly, two sets of experiments are performed. In the first set of experiments, some arbitrary scales (5, 10, 15, 20, 25, and 30) are chosen on which segmentation process is carried out. In the second set of experiments scales for different land covers are computed automatically by the proposed algorithm. It is found from the experimental results that the accuracies are better for the automated scale section. Moreover, the automatic scale estimation method has some remarkable advantages like (i) it reduces human effort for manually selecting proper scale for each land cover in the studied scene, (ii) the accuracies are higher as compared to established methods; (iii) in most of the cases, accuracy, sensitivity and specificity are also higher for automatic multi-scale RoIs selection method than manual process. In automatic scale selection method, segmentation is done only one time. Sometimes, the computed scales are close enough to the manually selected scales for some land covers and sometimes, they are dissimilar. The reason behind dissimilarity is that all the land covers are close to each other statistically. It is experimentally found that all the land covers are closer to each other statistically. Moreover, the dynamic ranges of intensities of different land covers are very close and sometimes they are overlapped. The the accuracies are satisfactory using the RF. Three state-of-the-art methods are adopted and are implemented on two datasets. Experimental results demonstrate that the proposed method outperforms three state-of-the-art methods quantitatively and qualitatively. Convolution neural networks (CNN) has made incredible improvements over the past few years. So, CNN would be used in near future for land covers classification.

**Acknowledgements** A. Seal is thankful to MeITY, Govt. of India for proving him Young Faculty Research Fellowship under Visvesvaraya PhD Scheme for Electronics and IT. This work has been partially funded by the Water Research Center For Agriculture and Mining, CRHIAM (CONICYT-FONDAP-1513001).

## References

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Susstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11), 2274–2282.
- Aplin, P., & Smith, G. M. (2008). Advances in object-based image classification. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37, 725–728.
- Baatz, M., & Shape, A. (2000). *Multiresolution segmentation-an optimization approach for high quality multi-scale image segmentation*. [http://www.agit.at/papers/2000/baatz\\_FP\\_12.pdf](http://www.agit.at/papers/2000/baatz_FP_12.pdf)
- Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3), 239–258.
- Bhattarai, N., Quackenbush, L. J., Calandra, L., Im, J., & Teale, S. (2011). Spectral analysis of scotch pine infested by *SirexNoctilio*. In *ASPRS annual conference, Milwaukee, Wisconsin*.

- Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2–16.
- Blaschke, T., Burnett, C., & Pekkarinen, A. (2004). Image segmentation methods for object-based analysis and classification. In S. M. De Jong & F. D. Van der Meer (Eds.), *Remote sensing image analysis: Including the spatial domain* (pp. 211–236). Dordrecht: Springer.
- Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., et al. (2010). Geographic object-based image analysis—Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 180–191.
- Blaschke, T., & Lang, S. (2006). Object based image analysis for automated information extraction—A synthesis. In *Measuring the Earth II ASPRS fall conference, San Antonio, Texas* (pp. 6–10).
- Blaschke, T., Lang, S., & Hay, G. (2008). *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*. Berlin: Springer.
- Blaschke, T., & Strobl, J. (2001). What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GeoBIT/GIS*, 6, 12–17.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140.
- Breiman, L. (2000). Some infinity theory for predictor ensembles. In *Technical Report 579. Statistics Department, UCB*.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Breiman, L. (2004). *Consistency for a simple model of random forests*. Technical Report 670. <https://www.stat.berkeley.edu/~breiman/RandomForests/consistencyRFA.pdf>. Accessed 6 Oct 2019.
- Breiman, L. (2017). *Classification and regression trees*. London: Routledge.
- Carleer, A. P., Debeir, O., & Wolff, E. (2005). Assessment of very high spatial resolution satellite image segmentations. *Photogrammetric Engineering & Remote Sensing*, 71(11), 1285–1294.
- Castilla, G., & Hay, G. J. (2008). Image objects and geographic objects. In T. Blaschke, S. Lang, & G. J. Hay (Eds.), *Object-based image analysis* (pp. 91–110). Berlin: Springer.
- Dey, V., Zhang, Y., & Zhong, M. (2010). A review on image segmentation techniques with remote sensing perspective. In *ISPRS TC VII symposium* (Vol. 38, pp. 31–42).
- Digital Globe. (2011). [http://worldview2.digitalglobe.com/docs/WorldView-2\\_8-Band\\_Applications\\_Whitepaper.pdf](http://worldview2.digitalglobe.com/docs/WorldView-2_8-Band_Applications_Whitepaper.pdf).
- Drăguț, L., Csillik, O., Eisank, C., & Tiede, D. (2014). Automated parameterisation for multiscale image segmentation on multiple layers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88, 119–127.
- Drăguț, L., Eisank, C., & Strasser, T. (2011). Local variance for multiscale analysis in geomorphometry. *Geomorphology*, 130(3), 162–172.
- Drăguț, L., Tiede, D., & Levick, S. R. (2010). ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science*, 24(6), 859–871.
- Dungan, J. L., Perry, J. N., Dale, M. R. T., Legendre, P., Citron-Pousty, S., Fortin, M. J., et al. (2002). A balanced view of scale in spatial statistical analysis. *Ecography*, 25, 626–640.
- Duro, D. C., Franklin, S. E., & Dube, M. G. (2012). A comparison of pixel based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing Environment*, 118(6), 259–272.
- Espindola, G. M., Camara, G., Reis, I. A., Bins, L. S., & Monteiro, A. M. (2006). Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. *International Journal of Remote Sensing*, 27(14), 3035–3040.
- Fisher, P. (1997). The pixel: A snare and a delusion. *International Journal of Remote Sensing*, 18(3), 679–685.
- Fukunaga, K., & Hostler, L. D. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1), 32–40.
- Gao, B. C. (1996). NDWI-A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3), 257–266.
- Gao, Y., & Mas, J. F. (2008). A comparison of the performance of pixel-based and object-based classifications over images with various spatial resolutions. *Online Journal of Earth Sciences*, 2(1), 27–35.
- Garcia-Pedrero, A., Gonzalo-Martin, C., Fonseca-Luengo, D., & Lillo-Saavedra, M. (2015). A GEOBIA methodology for fragmented agricultural landscapes. *Remote Sensing*, 7(1), 767–787.
- Geerken, R., Zaitchik, B., & Evans, J. P. (2005). Classifying rangeland vegetation type and coverage from NDVI time series using Fourier Filtered Cycle Similarity. *International Journal of Remote Sensing*, 26(24), 5535–5554.
- Geiß, C., Klotz, M., Schmitt, A., & Taubenböck, H. (2016). Object-based morphological profiles for classification of remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10), 5952–5963.

- Gholoobi, M., Tayyebi, A., Taleyi, M., & Tayyebi, A. H. (2010). Comparing pixel based and object based approaches in land use classification in mountainous areas. In *International archives of the photogrammetry, remote sensing and spatial information science, Kyoto, Japan* (Vol. 38).
- Gonzalo-Martín, C., Lillo-Saavedra, M., Gómez-Carpintero, J., García-Pedrero, Á. M., & Menasalvas, E. (2013). Unsupervised multi-scale ROIs determination for supervised thematic classification. *Towards Horizon*, 2020, 783–790.
- Hay, G. J., & Marceau, D. J. (2004). Multiscale object-specific analysis (MOSA): An integrative approach for multiscale landscape analysis. In S. M. De Jong & F. D. Van der Meer (Eds.), *Remote sensing image analysis: Including the spatial domain* (pp. 71–92). Dordrecht: Springer.
- Huang, X., & Zhang, L. (2008). An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 46(12), 4173–4185.
- Johnson, B., & Xie, Z. (2011). Unsupervised image segmentation evaluation and refinement using a multi-scale approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(4), 473–483.
- Karl, J. W., & Maurer, B. A. (2010). Spatial dependence of predictions from image segmentation: A variogram-based method to determine appropriate scales for producing land-management information. *Ecological Informatics*, 5(3), 194–202.
- Karlekar, A., Seal, A., Krejcar, O., & Gonzalo-Martín, C. (2019). Fuzzy k-means using non-linear s-distance. *IEEE Access*, 7, 55121–55131. <https://doi.org/10.1109/ACCESS.2019.2910195>.
- Kim M., Madden M., & Warner T. (2008) Estimation of optimal image object size for the segmentation of forest stands with multispectral IKONOS imagery. In T. Blaschke, S. Lang, & G. J. Hay (Eds.), *Object-based image analysis*. Lecture notes in geoinformation and cartography. Berlin, Heidelberg: Springer.
- Kurtz, C., Passata, N., Gańczarska, P., & Puissant, A. (2012). Extraction of complex patterns from multi-resolution remote sensing images: A hierarchical top-down methodology. *Pattern Recognition*, 45(2), 685–706.
- Meinel, G., & Neubert, M. (2004). A comparison of segmentation programs for high resolution remote sensing data. *International Archives of Photogrammetry and Remote Sensing*, 35, 1097–1105.
- Ming, D., Cai, H., Li, L., Qiao, C., & Du, J. (2012). Semivariogram-based spatial bandwidth selection for remote sensing image segmentation with mean-shift algorithm. *IEEE Geoscience and Remote Sensing Letters*, 9(5), 813–817.
- Ming, D., Li, J., Wang, J., & Zhang, M. (2015). Scale parameter selection by spatial statistics for GeOBIA: Using mean-shift based multi-scale segmentation as an example. *ISPRS Journal of Photogrammetry and Remote Sensing*, 106, 28–41.
- Ming, D., & Zhang, X. (2016). Cropland extraction based on OBIA and adaptive scale pre-estimation. *Photogrammetric Engineering and Remote Sensing*, 82(8), 635–664.
- Mohan, B. K., & Ladha, S. N. (2009). Comparison of object based and pixel based classification of high resolution satellite images using artificial neural networks. In *National symposium on security and soft computing*. Mumbai: IIT Bombay.
- Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.
- Quiles, M. G., Wang, D. L., Zhao, L., Romero, R. A. F., & Huang, D. S. (2011). Selecting salient objects in real scenes: An oscillatory correlation model. *Neural Networks*, 24, 54–64.
- Seal, A., Bhattacharjee, D., & Nasipuri, M. (2016). Human face recognition using random forest based fusion of à-trous wavelet transform coefficients from thermal and visible images. *AEU-International Journal of Electronics and Communications*, 70(8), 1041–1049.
- Seal, A., Bhattacharjee, D., Nasipuri, M., & Basu, D. K. (2015). UGC-JU face database and its benchmarking using linear regression classifier. *Multimedia Tools and Applications*, 74(9), 2913–2937.
- Shapiro, R., Freund, Y., Bartlett, P., & Lee, W. (1997). Boosting the margin: A new 395 explanation for the effectiveness of voting methods. *The Annals of Statistics*, 26(5), 1651–1686.
- Sharma, K. K., & Seal, A. (2019). Modeling uncertain data using Monte Carlo integration method for clustering. *Expert Systems with Applications*, 137, 100–116.
- Sheikh, Y. A., Khan, E. A., & Kanade, T. (2007). Mode-seeking by medoidshifts. In *2007 IEEE 11th international conference on computer vision* (pp. 1–8).
- Tan, P. N. (2018). *Introduction to data mining*. Noida: Pearson Education.
- Trias-Sanz, R., Stamon, G., & Louchet, J. (2008). Using colour, texture, and hierarchical segmentation for high-resolution remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(2), 156–168.
- UcaAvci, Z. D., Karaman, M., Ozelkan, E., & Papila, I. (2011). A comparison of pixel-based and object-based classification methods, a case study: Istanbul, Turkey. In *34th International symposium on remote sensing of environment, Sydney*.

- Vedaldi, A., & Soatto, S. (2008). Quick shift and kernel methods for mode seeking. ECCV, Part IV. In *European conference on computer vision*, Springer, Berlin, Heidelberg (pp. 705–718).
- Wang, M., Dong, Z., Cheng, Y., & Li, D. (2018). Optimal segmentation of high-resolution remote sensing image by combining superpixels with the minimum spanning tree. *IEEE Transactions on Geoscience and Remote Sensing*, 56(1), 228–238.
- Whiteside, T., & Ahmad, W. (2005). A comparison of object-oriented and pixel-based classification methods for mapping land cover in Northern Australia. In *Proceedings of SSC2005 spatial intelligence, innovation and praxis: The National Biennial Conference of the Spatial Sciences Institute*.
- Witharana, C., & Civco, D. L. (2014). Optimizing multi-resolution segmentation scale using empirical methods: Exploring the sensitivity of the supervised discrepancy measure Euclidean distance 2 (ED2). *ISPRS Journal of Photogrammetry and Remote Sensing*, 88, 108–121.
- Woodcock, C. E., & Strahler, A. H. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment*, 21(3), 311–332.
- Zhang, H., Fritts, J. E., & Goldman, S. A. (2008). Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110(2), 260–280.
- Zhou, Y., Li, J., Feng, L., Zhang, X., & Hu, X. (2017). Adaptive scale selection for multiscale segmentation of satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(8), 3641–3651.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Ayan Seal** received the PhD degree in engineering from Jadavpur University, West Bengal, India, in 2014. He is currently an Assistant Professor with the Computer Science and Engineering Department, PDPM Indian Institute of Information Technology, Design and Manufacturing Jabalpur, Madhya Pradesh, India. He has visited the Universidad Politécnica de Madrid, Spain as a visiting research scholar. He is the recipient of several awards. Recently, he has received Sir Visvesvaraya Young Faculty Research Fellowship from Media Lab Asia, Ministry of Electronics and Information Technology, Government of India. He has authored or co-authored of several journals, conferences and book chapters in the area of biometric and medical image processing. His current research interests include image processing and pattern recognition.



**Angel Garcia-Pedrero** received the B.Sc. degree in Computer Systems Engineering from the Villahermosa Institute of Technology (2006), the M.Sc. degree in Computer Science from the National Institute of Astrophysics, Optics and Electronics in Mexico (2009) and the Ph.D. degree in Advanced Computing for Science and Engineering from the Universidad Politécnica de Madrid (2016). He joined the Computer School at the Universidad Politécnica de Madrid in 2019, where he is currently Assistant Professor. His main research interests include image analysis, machine learning, and deep learning.



**Debotosh Bhattacharjee** is working as a full professor in the Department of Computer Science and Engineering, Jadavpur University with a fourteen years of post-PhD experience. His research interests pertain to the applications of machine learning techniques for Face Recognition, Gait Analysis, Hand Geometry Recognition, and Diagnostic Image Analysis. He has authored or coauthored more than 250 journals, conference publications, including several book chapters in the areas of Biometrics and Medical Image Processing. Two US patents have been granted on his works. Prof. Bhattacharjee has been granted sponsored projects by the Govt. of India funding agencies like Department of Biotechnology(DBT), Department of Electronics and Information Technology (DeitY), University Grants Commission(UGC) with a total amount of around INR 2 Crore. For postdoctoral research, Dr. Bhattacharjee has visited different universities abroad like the University of Twente, The Netherlands; Instituto Superior Técnico, Lisbon, Portugal; University of Bologna, Italy; ITMO National Research University, St. Petersburg, Russia; University of Ljubljana, Slovenia; Northumbria University, Newcastle Upon Tyne, UK and Heidelberg University, Germany. He is a life member of Indian Society for Technical Education (ISTE, New Delhi), Indian Unit for Pattern Recognition and Artificial Intelligence (IUPRAD), and a senior member of IEEE (USA).



**Mita Nasipuri** received her B.E.Tel.E., M.E.Tel.E., and Ph.D. (Engg.) degrees from Jadavpur University, in 1979, 1981 and 1990, respectively. Prof. Nasipuri has been a faculty member of J.U since 1987. Her current research interest includes image processing, pattern recognition, and multimedia systems. She is a senior member of the IEEE, U.S.A., Fellow of I.E (India) and W.B.A.S.T, Kolkata, India.



**Mario Lillo-Saavedra** received the B.E. degree in electrical engineering and the M.Sc. degree from the University of Concepción, Concepción, Chile, in 1997 and 2001, respectively, and the Ph.D. degree in computer sciences from the Technical University of Madrid, Madrid, Spain, in 2005. He is currently a Full Professor of University of Concepción. His current research interests include data fusion, machine learning, remote sensing applied to water management. He is a member of the Water Research Center for Agriculture and Mining (CHRIAM).



**Ernestina Menasalvas** Computer Scientist and PhD in Computer Science. She leads the MIDAS “Data Mining and data simulation group” at the Center of Biotechnology in UPM and she is databases and data mining professor at UPM. Her research integrates different aspects of data analytics; with the involvement in different real-world problems, including health, aviation, She has participated in more than 20 projects (H2020, FP7, EIT-Health). She has published more than 40 papers in journals including Data and Knowledge Engineering Journal, Physics Reports, Information Sciences, Expert Systems with applications and Journal of Medical Systems and International.



**Consuleo Gonzalo-Martin** received the Ph.D from the Complutense University of Madrid, in Physics, in 1989. She is currently a Full Professor at Universidad Politécnica de Madrid (Spain) Since September 2012, she is doing her research activities at the Center for Biomedical Technology of the UPM. Her main contributions have been in the areas of image processing and the development of new artificial neural networks algorithms in remote sensing, medical imaging and recognition of faces. At present, her main line of research is new methodologies for image analysis based on objects (OBIA) for satellite images and medical images and machine learning applied to the automatic image understanding. Also she is involved in text and image mining in the health care domain.