Sustain. Water Resour. Manag.
DOI 10.1007/s40899-017-0184-6

ORIGINAL ARTICLE

Application and comparison of advanced supervised classifiers in extraction of water bodies from remote sensing images

Arati Paul[1] · **Devarati Tripathi**[2] · **Dibyendu Dutta**[1]

Received: 18 November 2016 / Accepted: 23 August 2017

Abstract Water body extraction plays an important role in monitoring and assessing the existing water resources. It is a complex process that may be affected by many factors. This paper examines the major and advanced supervised classification approaches and ventures into the effectiveness of these techniques in extraction of water bodies from satellite images. The different classification techniques used for this purpose include support vector machine, artificial neural network, K-nearest neighbor, discriminant analysis and random forest. Commonly used normalized difference water index technique has also been examined in the study. Comparisons have been drawn among various variants of these methods and the accuracy in each case has been recorded. Each classification technique has been applied on input images from three different satellite sensors of varying spatial and spectral resolution, to compare their performance on different data sets of three different study areas. The study has found that supervised classifier can extract water bodies with a good accuracy from remotely sensed images even with a fewer number of labeled samples. Additionally, it is seen that the linear classifiers also yield good accuracy in extracting water bodies across different sensor's data.

Keywords Water bodies · Supervised · Classification · Remote sensing · Extraction

✉ Arati Paul
aratipaul@yahoo.com

[1] Regional Remote Sensing Centre – East, ISRO, Kolkata, India

[2] Department of Computer Science and Engineering, Institute of Engineering and Management, Kolkata, India

Introduction

Water bodies (including ocean) cover 70% of the earth's surface and play a fundamental role in the global carbon cycle and climate variations. They contribute in environmental and regional planning, transportation, industrial and agricultural production, etc. Over the years there has been an increase in water scarcity (Keller et al. 2000) due to water resources being subject to intense exploitation. Therefore, as an imperative to their sustainable management, there is a dire need to assess the extent and rate of their degradation and disappearance. The extraction of water bodies from a given remotely sensed image not only provides a platform to monitor and assess the existing water resources, but also develops a root for flood hazard prediction, flood frequency analysis (Dalrymple 1960), climate change effects, etc. Accurate mapping of surface water is essential for both research and policy making.

In the present study water body extraction is performed from satellite images. This is because remotely sensed images are a major source of consistent and continuous data for studying different types of land cover, water bodies, different features of the atmosphere, etc., at a variety of spatial and temporal scales. Moreover, they provide an easy, cost-effective and less time-consuming alternative for acquiring data. Measurements can be performed over large distances (several hundred or thousand kilometers in case of satellite sensors). Thus remote sensing provides an effective and efficient solution for acquiring data at large synoptic scales (Roughgarden et al. 1991; Barnsley 1999).

There are several available methods for extraction of water bodies from satellite images (Amarsaikhan and Douglas 2004). Unsupervised classification involves pixels being grouped into "clusters". The unsupervised image

Published online: 01 September 2017

classification technique is commonly used when no sample sites exist (Lu and Weng 2007). The spectral classes do not represent features on ground and it is time consuming to interpret them. Moreover, the spectral properties vary over time across the image. Unsupervised classification also provides low accuracy in classification and the spectral classes often overlap between different land features (Dayan et al. 1999). Water index method uses water indices like Normalized Difference Water Index (NDWI) and Modified Normalized Difference Water Index (MNDWI) (Lu and Weng 2007; Ji et al. 2009) as parameters for extraction of water bodies. They enhance the interested ground object and inhibit the others causing extraction to be easier (Yang et al. 2011). NDWI often confuses water with vegetation and soil and is not suitable for extraction of shallower parts of water. The two indices often require additional threshold value to extract water bodies which in turn is tough to standardize as it varies with the image (Xu 2005). In fractal characterization, classification is based on the texture of land forms to be classified. Water bodies tend to have a smooth texture compared to buildings and vegetation and this factor is used for their extraction. However, this method does not take spectral features into account and the results vary with image resolution (Mishra and Prasad 2014). An automated method for extracting rivers and lakes (AMERL) (Jiang et al. 2014) from Landsat imagery is a method which combines water indices and image processing techniques to identify water pixels (Kamavisdar et al. 2013). The method applied various thresholds for noise reduction, shadow removal and removal of small segments. Moreover, to set the internal and external markers for application of the watershed algorithm, prior knowledge of the image is required. Perceptron model is a kind of supervised classification which uses a linear predictor function, combining some characteristics of the feature vectors and summing it with weights to provide an input to the binary hard limit output function. To improve accuracy, the method implements single pixel removal and assigns different size parameters to the largest cluster of pixels misclassified as a water body. It applies thresholds to the feature vectors, in the reflectance values of the SWIR band and to the different size parameters. However, the thresholds are not standardized and moreover the usage of a linear predictor function is not very adept in classifying water bodies (Mishra and Prasad 2014). Single-band threshold method takes into account the fact that compared to land and soil, water reflects little energy in the near-infrared and mid-infrared bands. Threshold is used for water extraction. Although single-band threshold method is easy to model and simple to operate, it often incorporates more noise in the classified image compared to other methods and is not suitable for extraction of small and narrow water bodies (Yang et al. 2011). Inter-spectrum relation method uses the feature that water has reflectance in bands TM2 and TM3 greater than that in bands TM4 and TM5, i.e., TM2 + TM3 > TM4 + TM5. This feature is used to distinguish water bodies from shadow and unused land. This method gives relatively good accuracy in extraction on large as well as small water bodies (Yang et al. 2011). However, the method also extracts residential areas as water bodies. Spectral pattern analysis after reading the input image converts the digital number of each pixel into its corresponding reflectance value. It then determines modulation of the spectral reflectance curve for each pixel vector. From the acquired modulation values it determines which of them represent water bodies by visual checking. It then applies a variety of thresholds to the modulation values to extract water bodies (Duong 2012). The method fails because determining which modulation values determines water bodies is done by visual checking (Rundquist et al. 1987). The thresholds have not been standardized and hence will vary from image to image. Supervised classification (Barandela and Juarez 2002) uses image pixels representing regions of known, homogenous surface composition—training areas—to classify unknown pixels (Perumal and Bhaskaran 2010; Otukei and Blaschke 2010; Fern´andez-Delgado et al. 2014). Classification is the technique by which pixels of an image are grouped into specified categories, for example land cover features, to provide some meaning to them. The classification scheme and the structure of the classes depend on the needs of the user. Although classification required prior knowledge of the area to be classified, supervised classification has many advantages (Chen and Stow 2002). It provides high accuracy and represents informational classes that specify features on the ground. The basic method does not change with change of data, viz. images from different sensor. Moreover, the training sets can be reused if they do not change.

This study employs supervised classification for extraction of water bodies of three different study areas from satellite images of three different sensors (Althausen 2002), namely LISS-III, AWIFS and LANDSAT 8 for each area. Five different classifier including advanced methods of supervised classification, viz. support vector machine (SVM), artificial neural network (ANN), random forest (RF), etc. (McFeeters 1996; Atkinson and Tatnall 1997), along with their variants have been applied in the present study. Commonly used NDWI threshold have also been examined for comparison. A comparative study among different methods has been made after measuring the overall accuracy and kappa's coefficient of each method. In the present study, it is seen that supervised classifier is capable of extracting surface water bodies of any size and shape efficiently in different types of input images.

Study area

The present study has been carried out in three different study areas geographically located in three different states in India (Fig. 1a). The brief description of each study area is provided in following sections.

Study area I

The first study area chosen is located in Masanjore district of Jharkhand, India, along the latitude 24°09′51.33″N and the longitude 87°18′56.42″E. The area consists of a dam, called the Masanjore Dam along with a river, the Mayurakshi River. In addition, the area also consists of water reservoirs like Digol Bandh, Bada Pokhra and several more in Dumba, Jharkhand. Various small tributaries of river Mayurakshi present in Karantola and neighboring areas are also present.

Masanjore experiences the cold-weather season from November to February. The hot weather season lasts from March to mid-June with May being the hottest season followed by monsoon lasting from mid-June to October. The location of the study area is depicted in Fig. 1b.

Study area II

The second study area chosen is located in Hingoli district of Maharashtra, India, along the latitude 19°37′12.0″N and the longitude 76°55′12.0″E. The area consists of a dam called the Siddheshwar Dam which is an earth fill dam built on Purna River near Siddheshwar.

Siddheshwar experiences summer season from March to May. The monsoons prevail from June to September with July being the wettest month, followed by the winter season spanning from November to February. The location of the study area is depicted in Fig. 1c.

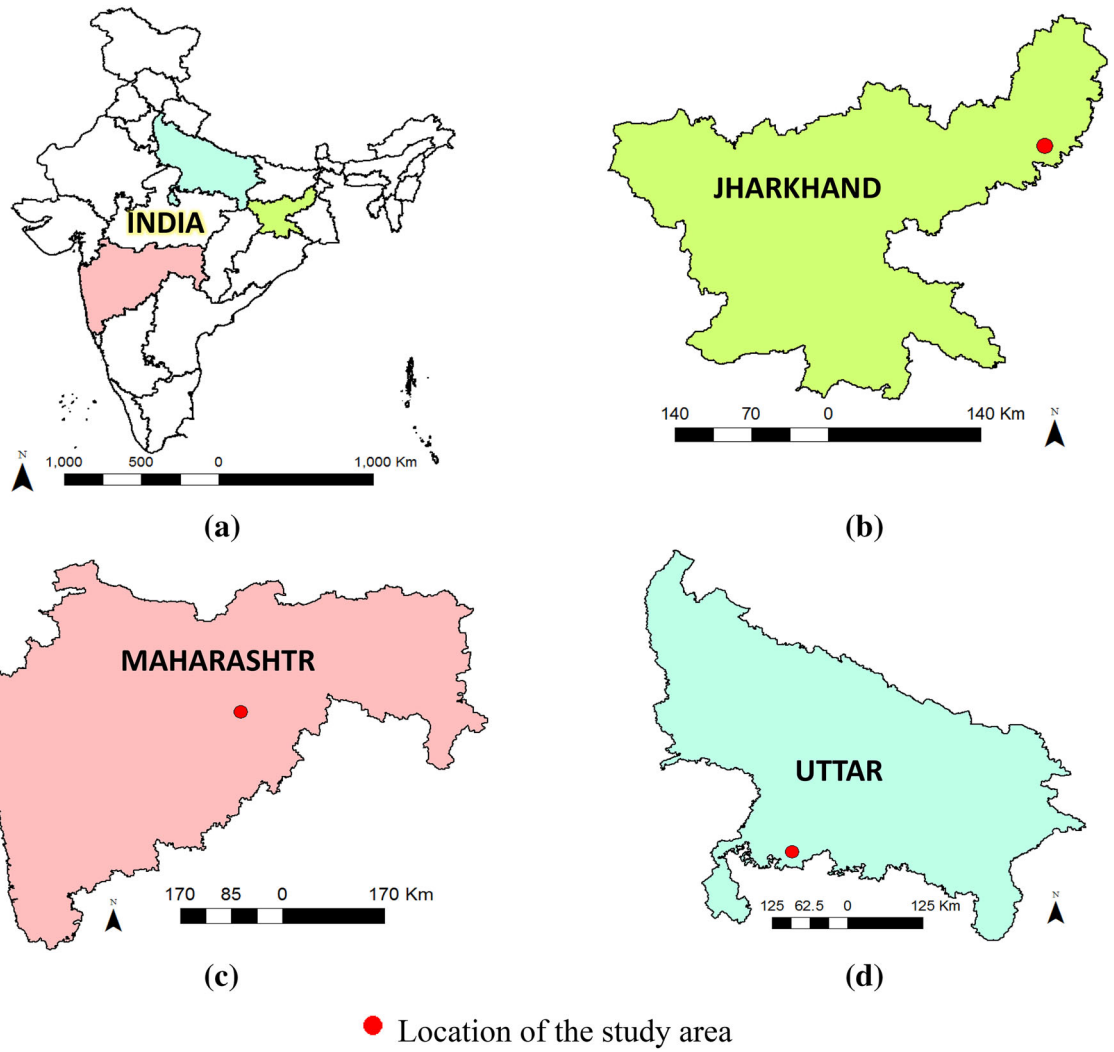

Fig. 1 **a** Map of India. Location of study area: **b** I, **c** II, and **d** III

Study area III

The third study area chosen is located in Mahoba district of Uttar Pradesh, India, along the latitude 25°22′48.0″N and the longitude 79°40′48.0″E. The area consists of a reservoir called the Arjun Sagar Reservoir and is connected to Arjun River.

Mahoba has the tropical monsoon type climate with October–February being the winter season, March–May the summer season and June–September, the rainy season. The location of the study area is depicted in Fig. 1d.

Data used

Water body extraction is performed using three kinds of satellite images of varied spatial and spectral resolution.

LISS-III (linear imaging self-scanner) data

The multispectral LISS-III image for the first study area is dated 25 November 2011(path 108/row 56) and has been acquired from IRS-Resourcesat-1 (www.bhuvan.nrsc.gov.in). Similarly the LISS-III images acquired on 23 November 2011(path 98/row 59) and 14 October 2008 (path 99/row 54) are used for the second and the third study area, respectively.

The image operates with three spectral bands in visible and near-infrared (VNIR) and one band in short-wave infrared (SWIR) with 23.5 m spatial resolution and a swath of 141 km. It has been orthorectified and cropped according to the study area.

AWIFS (advanced wide field sensor) data

The multispectral AWIFS images dated 20 January 2009 (Path 109/Row 58), 23 December 2013(path 100/row 59) and 22 November 2013(path 100/row 54), acquired from IRS-Resourcesat-2 (www.bhuvan.nrsc.gov.in), have been used for the first, second and third study area, respectively.

The AWIFS image operates with three spectral bands in visible and near-infrared (VNIR) and one band in short-wave infrared (SWIR) with 56 m spatial resolution and a combined swath of 730 km achieved through two AWiFS cameras. It has been orthorectified and cropped according to the study area.

LANDSAT data

The multispectral LANDSAT image for the first study area is dated 26 January 2016 (path 139/row 43) and has been acquired from Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) [www.glovis.usgs.gov]. The date of acquisition of the LANDSAT image for the second and the third study area is 14 March 2017 (path 145/row 46) and 08 April 2017 (path 144/row 42), respectively.

The image operates with seven spectral bands namely blue (0.45–5.2 μm), green (0.52–0.60 μm), red (0.63–0.69 μm), near-infrared (NIR-0.76–0.90 μm), and short-wave infrared 1(SWIR 1—1.55–1.75 μm) and short-wave infrared 2(SWIR 2—2.11–2.29 μm). The scene has been cropped according to the study area.

NDWI is often considered as a good indicator of water. Hence the NDWI (Xu 2006; Zhang et al. 2003) has been derived and added as another layer with the existing LANDSAT image and then classified. This exercise is performed to see the influence of NDWI layer as an additional information in supervised classification.

Methodology

Water body extraction from satellite image using supervised classifier is essentially a two-class classification problem that divides the image into two classes: water and non-water. A satellite image contains different types of land cover feature with different spectral signature patterns. After preparing all the data sets as described in section “Data used”, the location and the label of selected spectral signatures for each study area have been sampled out by visual interpretation of images and are saved in the form of shape files. A software tool, developed in MATLAB, includes different classifiers as well as accuracy assessment indices to perform classification and evaluate the output automatically. The location of labeled samples (the shape file) and the corresponding satellite image are supplied to the software tool that extracts the labeled patterns and divides them into training and testing sets. The training samples are used to train a selected classifier and testing samples are used to assess the classified image output. Detailed methodology is presented in following sections.

Creation of the training and testing set

The selected representative labeled samples are randomly divided into the training and testing samples based on a given percentage. The testing set is later used for accuracy analysis, while the training set is used for classification. To see the effectiveness of the supervised classifier in extracting water bodies, a very low percentage of total labeled sample, e.g., 0.1, 1.5 and 0.6% for LISS-III, AWIFS and LANDSAT image, respectively (for study area I) is used for training and the rest is used in testing. The

Table 1 Number of training samples used to train the classifier in study area I

Image	Water sample (in number of pixels)		Non-water sample (in number of pixels)	
	Training	Testing	Training	Testing
LISS-III	12	5244	22	41,367
AWiFS	15	1155	15	890
LANDSAT	11	1342	18	3500

Table 2 Number of training samples used to train the classifier in study area II

Image	Water sample (in number of pixels)		Non-water sample (in number of pixels)	
	Training	Testing	Training	Testing
LISS-III	13	261	21	516
AWiFS	14	59	18	117
LANDSAT	20	723	21	561

Table 3 Number of training samples used to train the classifier in study area III

Image	Water sample (in number of pixels)		Non-water sample (in number of pixels)	
	Training	Testing	Training	Testing
LISS-III	10	115	23	167
AWiFS	10	20	24	80
LANDSAT	10	60	53	494

pixels used for training have been uniformly distributed throughout the image. Tables 1, 2 and 3 summarize the details of training and testing samples and Fig. 2 depicts the spatial distribution of training and testing samples for each of the three different study area (Foody and Mathur 2004).

Performing the classification

After the creation of training set, classification is performed on the input image. Following classifiers are used in the present study.

Discriminant analysis

Discriminant analysis (DA) (Press and Wilson 1978; Johnson and Wichern 2007), also referred to as maximum likelihood classification in the remote sensing literature (Lillesand and Kiefer 1994; Jensen 2000), is a parametric classification technique that distinguishes classes by estimating multidimensional distances among classes (Hogland et al. 2013). In DA, new observations are classified into one of the known groups or clusters or populations based on some measured characteristics. In the generic method, we apply Bayes' rule and we determine the conditional probability that an observation came from the new population given the observed values of the multivariate vector of variables. The decision rule is provided in Eq. (1):

$$\text{Put } X \text{ in } i\text{th class, if } P_i * p_i(X) > = P_j * p_j(X) \text{ for all } j \neq i, \quad (1)$$

where P_i is the prior probability of class i, X is the observed pixel vector and $p_i(X)$ is conditional probability to observe X from class i, or probability density function (PDF).

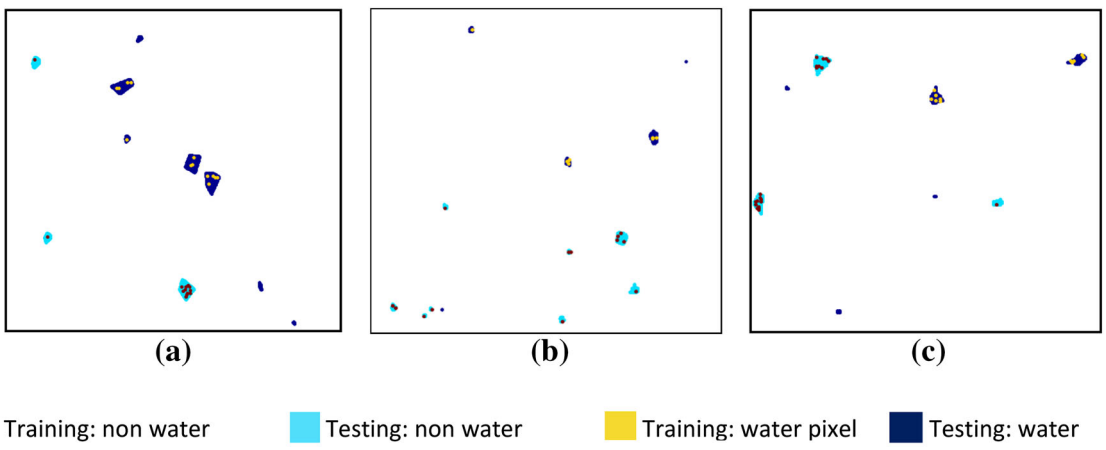

Fig. 2 **a–c** Spatial distribution of training and testing samples for study area I, II and III, respectively

In this study, we assume that this conditional probability follows multivariate Gaussian distribution (Cover and Hart 1967; Chiarandini 2013). Various types of discriminant analysis have been implemented for classification, namely, 'linear' which estimates one covariance matrix for all classes, quadratic which estimates one covariance matrix for each class, diagLinear which uses the diagonal of the 'linear' covariance matrix, and its pseudo-inverse if necessary, diagQuadratic which uses the diagonals of the quadratic covariance matrices, and their pseudo-inverses if necessary (Du and Chang 2001; Hastie and Tibshirani 1996).

K-nearest neighbor

In KNN algorithm, "K" (a positive integer) signifies the number of nearest neighbors. Then it finds K-nearest neighbors of observed pixel vector among all training points. Here the distance function utilized is the Euclidean distance function (Cover and Hart 1967). The observed pixel is assigned to that class where maximum number of nearest neighbor falls. Here no PDF and prior probability is required, but finding the appropriate "K" is subjective. The decision rule is provided in Eq. (2):

$$\text{Put } X \text{ in } i\text{th class}, \quad \text{if } K_i > = K_j \text{ for all } j \neq i, \tag{2}$$

where K_i is the number of nearest neighbor that belongs to the ith class and X is the observed pixel vector.

Support vector machine

A support vector machine (SVM) performs classification by finding the hyperplane that maximizes the margin between the two classes and minimizes the empirical risk. The vectors (cases) that define the hyperplane are the support vectors (Cortes and Vapnik 1995; Ben-Hur et al. 2001). Apart from linear classification, SVMs can also perform a non-linear classification using what is called the kernel trick by implicitly mapping their inputs into high-dimensional feature spaces. A variety of support vector machine algorithms have been used in the classification process namely, C-support vector classification and ν-support vector classification along with different types of kernel, viz. linear function, polynomial function and radial basis function (RBF) (Chang and Lin 2013; Foody et al. 1995).

The kernel on two samples x and x' represented as feature vectors in the input space.

The linear kernel is defined as (3):

$$K(x, x') = \langle x, x' \rangle. \tag{3}$$

The polynomial kernel is defined as (4):

$$K(x, x') = (\langle x, x' \rangle + 1)^d, \tag{4}$$

where d is the degree of polynomial.

The RBF kernel is defined as (5):

$$K(x, x') = \exp\left(\frac{||x - x'||^2}{2\sigma^2}\right), \tag{5}$$

where $||x - x'||^2$ may be recognized as the squared Euclidean distance between the two feature vectors and σ is a free parameter. An equivalent, but simpler, definition involves a parameter: $\gamma = -(1/2\sigma^2)$, provided in (6):

$$K(x, x') = \exp(\gamma ||x - x'||^2). \tag{6}$$

Artificial neural network

The artificial neutral network (ANN) is a system based on the biological neural network, such as the brain (McCulloch and Pitts 1943). An ANN comprises a network of artificial neurons (also known as "nodes"). These nodes are connected to each other, and the strength of their connections to one another is assigned a value based on their weightage (Farley and Clark 1954). If the value of the connection is high, then it indicates that there is a strong connection. Within each node's design, a transfer function is built in (Entezari-Maleki et al. 2009). In present study, the sigmoid transfer function is used. The number of hidden layers has been altered correspondingly to determine the best possible result (Erbek et al. 2004).

In an ANN the number of neuron in the input and the output layers are same as the number of features and the number of classes, respectively. The number of neurons chosen for the hidden layer is problem dependent. The state of the neurons present in the hidden or output layers re derived as (7):

$$I_j = \sum_{i=1}^{I} O_i W_{ji}, \tag{7}$$

where I_j is the net input (I) to the neuron j, O_i is the output of the ith neuron of the previous layer, W_{ji} is the weight value of the connection between neuron i and neuron j. The output of neuron j is obtained as: $O_j = f(Ij)$. f is the activation function and is determined using a sigmoid function as (8):

$$f(Ij) = \frac{1}{1 + \exp^{-I_j}}. \tag{8}$$

The network is trained to modify its weight through iterations to minimize the output error using labeled input samples.

Random forest

Random forest classification technique is an ensemble learning method for classification that operates by

constructing a multitude of decision trees (Friedl and Brodley 1997) at training time and outputting the class that is the mode of the classes of the individual trees (Ho 1995, 1998). Tree bagging takes place at the training stage to sample the training examples and train the decision tree. Tree bagging is followed by feature bagging where a random subset of features is selected at each candidate split. This allows the trees to be correlated or de-correlated (Breiman 2001; Lu and Weng 2007).

Consider N as the number of trees in the forest and M the number of variables in the random subset at each node. The RF algorithm works as follows (Liaw and Wiener 2002):

1. Draw N bootstrap samples from the labeled data.
2. For each of the bootstrap samples, grow an unpruned classification tree, with the following criterion: at each node, rather than choosing the best split among all predictors, randomly sample M number of the predictors and choose the best split from among those variables.
3. Predict the class of the observed pixel by taking the mode of predictions of N trees.

Normalized Difference Water Index (NDWI)

NDWI is calculated using Eq. (9):

$$\text{NDWI} = (\rho 2 - \rho 4)/(\rho 2 + \rho 4), \tag{9}$$

where ρ2 represents the reflectance in the green band and ρ4 represents the reflectance in the near-infrared (NIR) band. Threshold values are applied to the NDWI so as to eliminate the non-water surfaces which may have low reflectance, to distinguish them from the water surfaces. The threshold values for different images have been selected using Otsu's method (Otsu 1979).

Performance evaluation criteria

The test samples (created in section "Creation of the training and testing set") are used as the reference data for evaluation of performance. The classification of the new observation is compared with the reference data obtained from the testing set which accurately reflects the true class label. The accuracy measurement is performed using overall accuracy (OA) and kappa's coefficient (κ) (Fitzgerald and Lees 1993).

Overall accuracy

Overall accuracy (OA) is the ratio of the number of samples correctly classified to the total number of test samples. OA is measured from the confusion matrix $K_{C\times C}$, where C represents the number of classes. In present study, the value of C is 2. The element k_{ij} of the matrix K denotes the number of samples of the jth class which are classified into the ith class (for all i, j belonging to C). The overall accuracy (OA) is calculated using Eq. 10:

$$\text{OA} = \frac{\sum_{i=1}^{C} k_{ii}}{n}, \tag{10}$$

where n is the total number of test samples.

Kappa coefficient

The kappa coefficient (Congalton and Green 2009) is a measure defined as the difference between the actual

Table 4 Performance analysis for LISS-III image (dimensions: 1153*1153*4)

Classifier	Type		Time (ms)	OA	κ
Discriminant	Linear		13.89	**99.24**	99.97
	DiagLinear		8.50	99.07	99.93
	Quadratic		15.84	99.02	99.94
	DiagQuadratic		13.31	98.87	99.97
	Mahalanobis		15.75	99.81	99.99
SVM	SVM type	Kernel type			
	CSVC	Linear	14.18	99.27	99.94
		Polynomial	9.027	99.23	99.93
		Radial basis	15.42	99.23	99.98
KNN	Nearest neighbor: 10		22.21	88.73	99.99
RF	Decision trees: 5		54.64	96.56	99.95
ANN	Hidden layers: 10		40.63	99.28	99.93
NDWI	Threshold: 0.0344		1.10	98.75	99.71

Table 5 Performance analysis for AWIFS image (dimensions: 621*729*4)

Classifier	Type		Time (ms)	OA	κ
Discriminant	Linear		4.94	**99.98**	99.98
	DiagLinear		7.14	98.06	99.95
	Quadratic		7.03	99.70	99.93
	DiagQuadratic		7.01	96.67	99.90
	Mahalanobis		5.30	99.45	99.90
SVM	SVM type	Kernel type			
	CSVC	Linear	4.38	99.95	99.90
		Polynomial	7.29	99.31	99.90
		Radial basis	9.54	98.06	99.92
KNN	Nearest neighbor: 10		9.64	98.46	99.90
RF	Decision trees: 5		40.54	98.31	99.90
ANN	Hidden layers: 10		12.53	99.75	99.90
NDWI	Threshold: −0.0659		1.38	99.51	99.10

Table 6 Performance analysis for LANDSAT image (dimensions: 707*853*7)

Classifier	Type		Time (ms)	OA	κ
Discriminant	Linear		8.63	97.44	99.96
	DiagLinear		7.85	99.22	99.97
	Quadratic		8.11	94.88	99.96
	DiagQuadratic		8.48	98.22	99.96
	Mahalanobis		8.10	94.86	99.96
SVM	SVM type	Kernel type			
	CSVC	Linear	7.19	99.48	99.97
		Polynomial	7.02	99.44	99.96
		Radial basis	11.49	99.45	99.97
KNN	Nearest neighbor: 10		11.92	**99.55**	99.96
RF	Decision trees: 5		44.69	86.08	99.96
ANN	Hidden layers: 10		26.53	98.53	99.96
NDWI	Threshold: 0.0532		2.71	95.45	99.85

Table 7 Performance analysis for LISS-III image (dimensions: 546*654*4)

Classifier	Type		Time (ms)	OA	κ
Discriminant	Linear		2.08	97.81	99.77
	DiagLinear		2.13	97.68	99.79
	Quadratic		2.13	99.80	99.76
	DiagQuadratic		2.22	99.73	99.76
	Mahalanobis		2.18	**99.85**	99.85
SVM	SVM type	Kernel type			
	CSVC	Linear	1.99	99.83	99.76
		Polynomial	2.05	99.76	99.76
		Radial basis	2.89	99.52	99.74
KNN	Nearest neighbor: 10		2.79	98.42	98.97
RF	Decision trees: 5		10.13	99.34	99.75
ANN	Hidden layers: 10		3.46	99.47	99.76
NDWI	Threshold: 0.2288		1.01	99.25	99.39

Table 8 Performance analysis for AWIFS image (dimensions: 273*328*4)

Classifier	Type		Time (ms)	OA	κ
Discriminant	Linear		0.93	99.45	98.97
	DiagLinear		1.03	99.75	98.97
	Quadratic		0.97	98.85	98.99
	DiagQuadratic		0.87	98.85	98.99
	Mahalanobis		0.93	98.85	98.99
SVM	SVM type	Kernel type			
	CSVC	Linear	1.35	**99.87**	98.97
		Polynomial	1.02	98.88	98.97
		Radial basis	1.12	95.33	99.92
KNN	Nearest neighbor: 10		1.11	98.25	98.97
RF	Decision trees: 5		3.21	94.25	99.06
ANN	Hidden layers: 10		1.47	98.43	98.98
NDWI	Threshold: −0.0719		1.82	99.22	98.85

Table 9 Performance analysis for LANDSAT image (dimensions: 1056*1131*8)

Classifier	Type		Time (ms)	OA	κ
Discriminant	Linear		8.59	99.32	99.84
	DiagLinear		7.85	**99.77**	99.84
	Quadratic		6.38	94.54	99.85
	DiagQuadratic		8.31	99.06	99.84
	Mahalanobis		8.57	94.46	99.85
SVM	SVM type	Kernel type			
	CSVC	Linear	8.51	99.33	99.84
		Polynomial	8.37	99.57	99.84
		Radial basis	8.90	98.52	99.76
KNN	Nearest neighbor: 10		12.21	96.72	99.85
RF	Decision trees: 5		33.56	99.48	99.84
ANN	Hidden layers: 10		20.07	99.65	99.84
NDWI	Threshold: 0.0865		2.72	98.86	99.80

agreement in the confusion matrix and the chance agreement, which is indicated by row and column totals. The kappa coefficient is widely adopted, as it also takes into consideration the off-diagonal elements of the confusion matrix that are ignored in OA. The value of the kappa coefficient (κ) lies in the range 0 and 1. The κ value closer to 1 signifies better classification.

The kappa coefficient is calculated using Eq. (11):

$$\kappa = \frac{n\sum_{i=1}^{C} k_{ii} - \sum_{i=1}^{C} k_{i+}.k_{+i}}{n^2 - \sum_{i=1}^{C} k_{i+}.k_{+i}}, \tag{11}$$

where n is the total number of test samples. k_{i+} and k_{+j} represent sum of the elements of the ith row and jth column, respectively, in the confusion matrix.

Results and discussion

Each input image is classified using each of the classifiers including their variants (as discussed in previous section) separately and the performance of each is evaluated. A detailed assessment of these methods with respect to their accuracy and execution time requirement is given in Tables 4, 5 and 6 for LISS-III, AWIFS and LANDSAT image, respectively, for the study area I. It is observed from the performance analysis tables (Tables 4, 5, 6) that for the LISS-III, AWIFS and LANDSAT images highest accuracy (marked in Bold) is obtained when Mahalanobis discriminant, linear discriminant and KNN classification is performed, respectively. For the second study area

Table 10 Performance analysis for LISS-III image (dimensions: 251*378*4)

Classifier	Type		Time (ms)	OA	κ
Discriminant	Linear		1.09	99.15	99.33
	DiagLinear		0.91	99.31	99.45
	Quadratic		1.34	97.91	99.35
	DiagQuadratic		0.89	99.44	99.33
	Mahalanobis		0.89	97.56	99.35
SVM	SVM type	Kernel type			
	CSVC	Linear	0.90	99.21	99.33
		Polynomial	0.88	99.50	99.78
		Radial basis	1.00	98.07	99.46
KNN	Nearest neighbor: 10		1.17	99.23	99.87
RF	Decision trees: 5		4.71	**99.52**	99.33
ANN	Hidden layers: 10		6.09	99.41	99.33
NDWI	Threshold: −0.0459		1.06	98.10	99.90

Table 11 Performance analysis for AWIFS image (dimensions: 190*258*4)

Classifier	Type		Time (ms)	OA	κ
Discriminant	Linear		1.27	99.21	99.31
	DiagLinear		1.13	99.13	99.46
	Quadratic		1.23	99.29	99.32
	DiagQuadratic		1.11	98.86	98.99
	Mahalanobis		1.14	99.29	99.31
SVM	SVM type	Kernel type			
	CSVC	Linear	1.47	99.10	99.31
		Polynomial	0.85	98.88	99.78
		Radial basis	0.98	97.93	99.46
KNN	Nearest neighbor: 10		1.39	98.65	99.87
RF	Decision trees: 5		5.12	**99.30**	99.62
ANN	Hidden layers: 10		8.65	99.22	98.98
NDWI	Threshold: −0.0093		1.03	98.29	99.90

(Tables 7, 8, 9), highest OA% (marked in Bold) is obtained for Mahalanobis discriminant, CSVC and diagLinear discriminant classifier for LISS-III, AWIFS and LANDSAT images, respectively. Similarly for the third study area (Tables 10, 11, 12), RF provided the maximum OA% for LISS-III and AWIFS while the best OA% (marked in Bold) is obtained from diagLinear discriminant classifier for the

Table 12 Performance analysis for LANDSAT image (dimensions: 253*307*7)

Classifier	Type		Time (ms)	OA	κ
Discriminant	Linear		3.81	99.28	99.77
	DiagLinear		1.14	**99.55**	99.78
	Quadratic		1.80	98.19	99.78
	DiagQuadratic		0.94	98.91	99.78
	Mahalanobis		0.90	98.01	99.79
SVM	SVM type	Kernel type			
	CSVC	Linear	1.67	99.28	99.77
		Polynomial	0.85	99.27	99.78
		Radial basis	1.28	98.01	99.82
KNN	Nearest neighbor: 10		2.47	98.37	99.78
RF	Decision trees: 5		6.77	99.09	99.78
ANN	Hidden layers: 10		17.30	98.55	99.77
NDWI	Threshold: −0.0879		1.23	99.53	99.95

LANDSAT image. Hence it is seen that supervised classifiers extract water bodies with good accuracy.

The diagQuadratic discriminant classifier for AWIFS image in study area II and CSVC-polynomial for all three data sets in study area III requires the shortest amount of time, while NDWI method takes the minimum amount of time for rest (6 out of 9 number) of the data sets, which can be attributed to the nature of the method (Tables 4, 5, 6, 7, 8, 9, 10, 11 and 12).

The relative performance with respect to overall accuracy is depicted in Fig. 3 for each algorithm and for each study area. From Fig. 3 it is seen that the overall performance of all the tested methods in extracting water bodies from remote sensing image is good. The effect of spatial resolution of input image for water body extraction using these methods is negligible as the relative OA% change of three different types of images is random in nature for each method. Hence supervised classification as well as NDWI methods is robust while considering different spatial resolution of input data.

It is also seen from Fig. 3 that instead of enhancing the feature information (band) by introducing the NDWI layer in LANDSAT image the outcome does not get affected much. Rather the performance is comparable with other images where NDWI is not considered. Although supervised classification has a potential to accommodate additional information in terms of feature layer, here it did

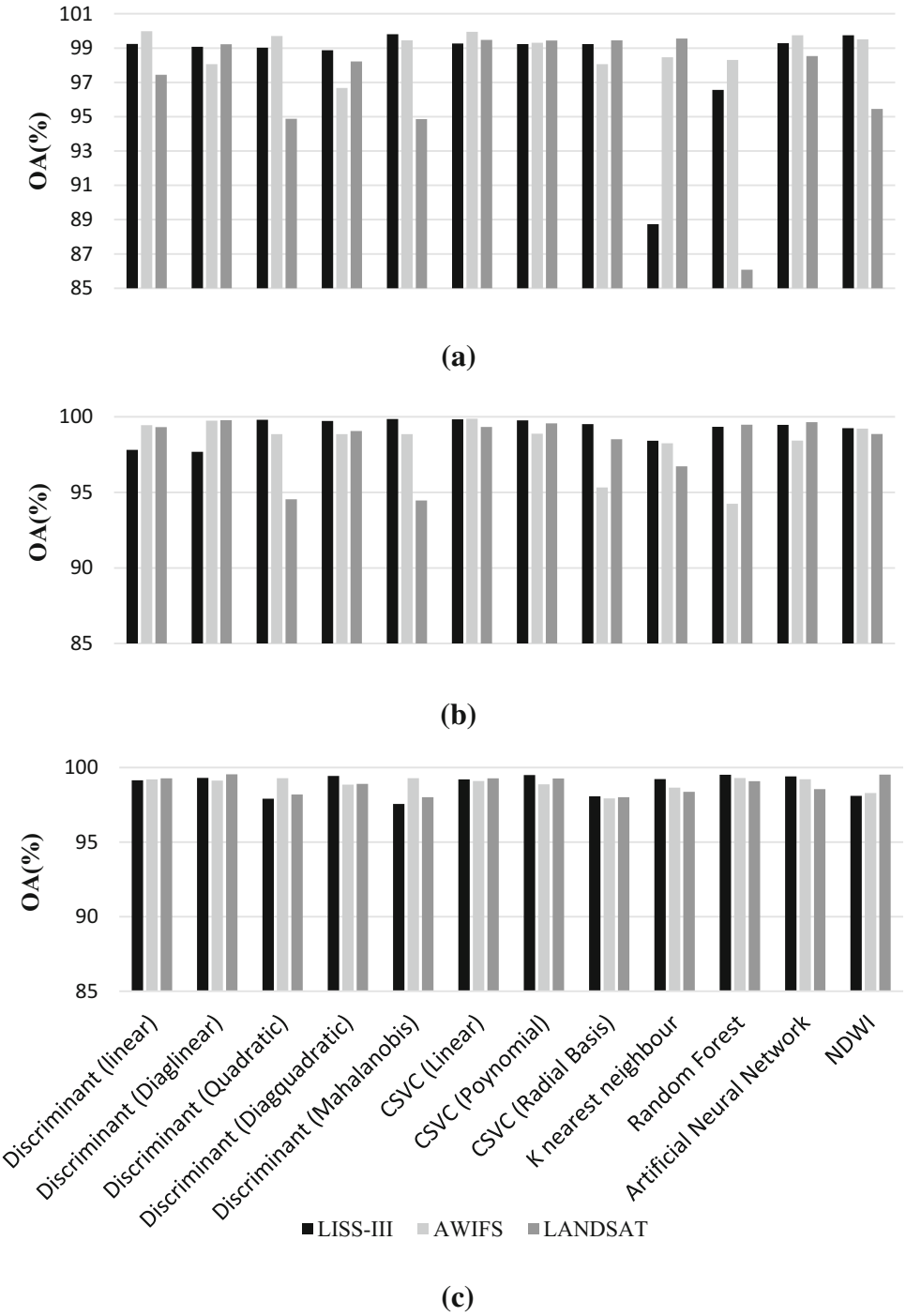

Fig. 3 **a**, **b**, **c** shows overall accuracies of different classification methods for LISS-III, AWIFS and LANDSAT images of study area I, II and III, respectively

not affect the result because the additional layer is a derivative of the few other existing layers. The information content of the additional feature layer would have been enriched if it would have been generated from different source or other observations.

The process was executed on a 64-bit machine with a 2.3 GHz Intel Core i5-4200U processor having 5.89 GB RAM. Figure 4 represents the relative execution time taken for each algorithm for all images for each study area. From Fig. 4 it is seen that execution time taken by any of the classifiers is less than a second (highest value 54.64 ms). Hence it can be concluded that all the classifiers are computationally effective. It is also seen in the figure that the execution time is directly proportional to the dimension

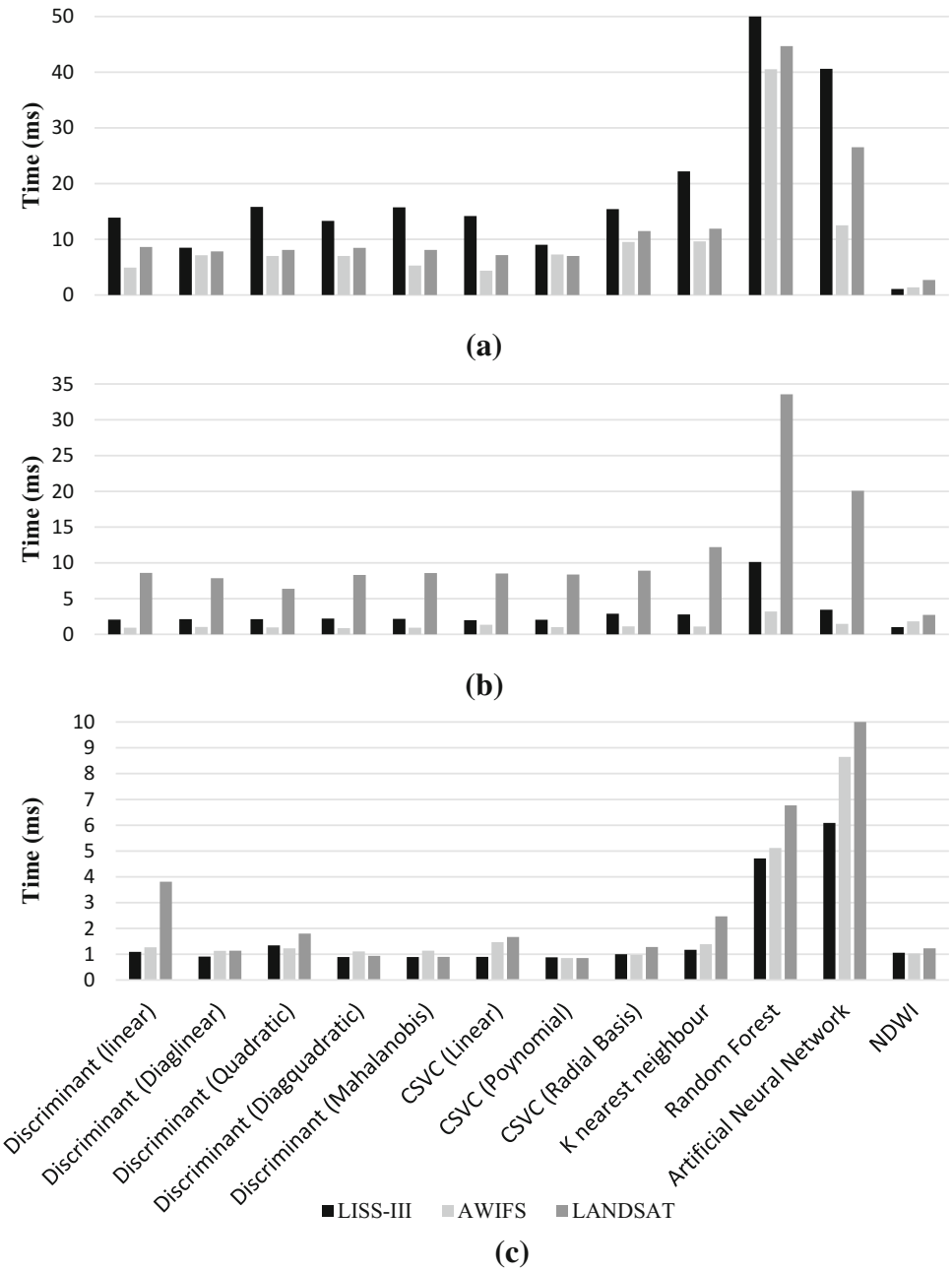

Fig. 4 **a**, **b**, **c** shows execution time required by different classification methods for LISS-III, AWIFS and LANDSAT images of study area I, II and III, respectively

of the input data. Among all classifiers RF and ANN have taken more time which may be attributed to the nature of training their respective classifier.

The output images containing extracted water bodies from LISS-III, AWIFS and LANDSAT data by using each of the above-mentioned classifiers are shown in Figs. 5, 6 and 7 for study areas I, II, III, respectively. It is seen from the figures (number 5–7) that the supervised classifiers are capable of extracting water bodies of different shapes as well as sizes. If required, size threshold can also be applied on the output to segregate smaller and larger water bodies. Streams are well extracted in LANDSAT image compared to other two images by almost all classifiers (Fig. 5), which can be ascribed to variations in training samples as well as spectral variation of data sets. As observed in Figs. 5, 6,

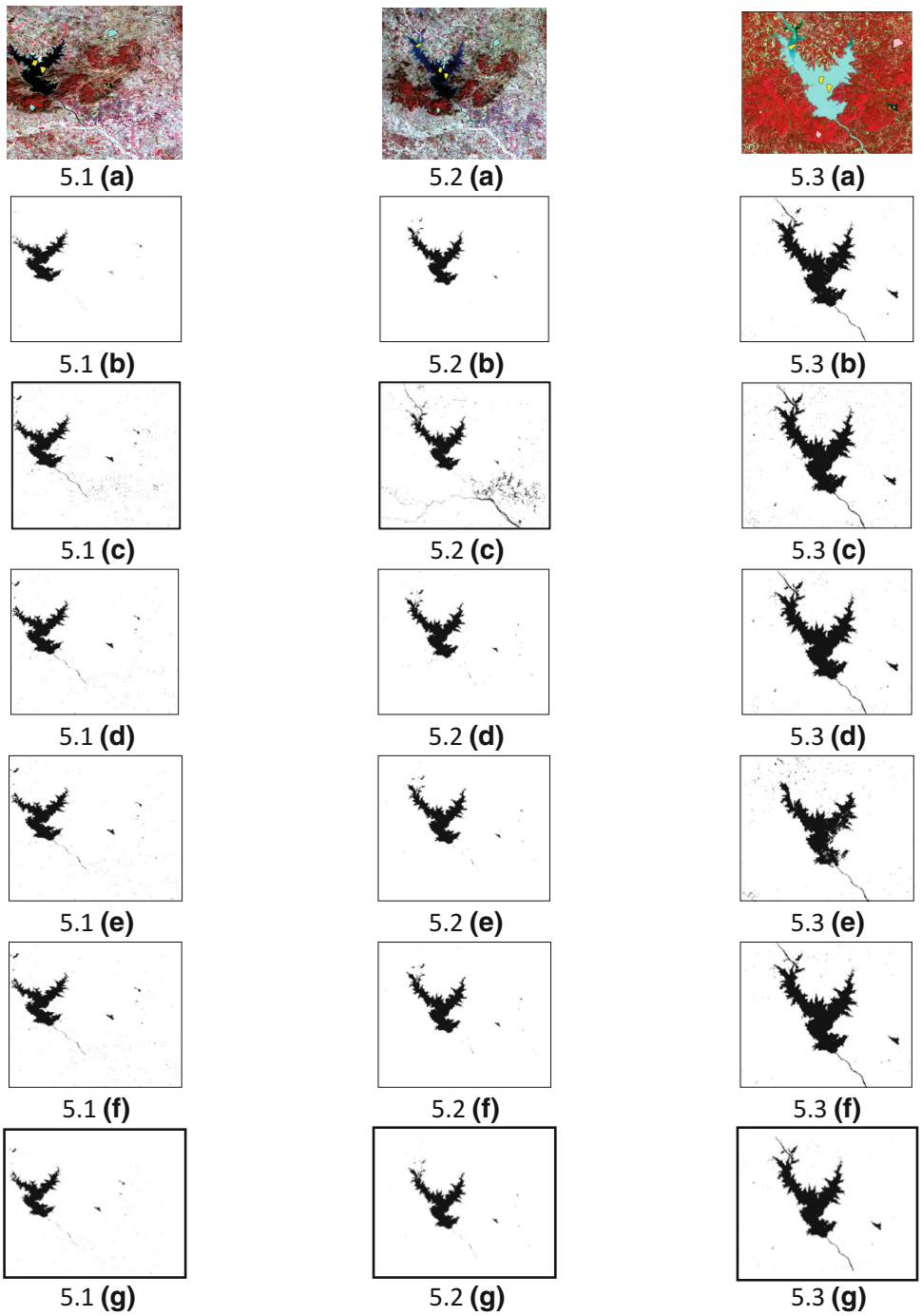

Fig. 5 Study area I: each column (*left* to *right*) represents **a** false color composite (FCC) of LISS-III, AWiFS and LANDSAT data (overlaid with labeled samples) along with extracted water bodies using: **b** Mahalanobis discriminant, **c** CSVC-polynomial, **d** KNN, **e** RF, **f** ANN and **g** NDWI methods, respectively

7, SVM polynomial, RF (in Fig. 61.e), KNN (in Fig. 73. d) and ANN (in Fig. 73.f) extracted water bodies in detail compared to other classifiers. Through visual inspection it is also seen that dam pixels are identified correctly in most of the output of supervised classifiers even if the training samples were not provided for those cases. But the same is not witnessed in the thresholded output of the NDWI. Although the method is simple and

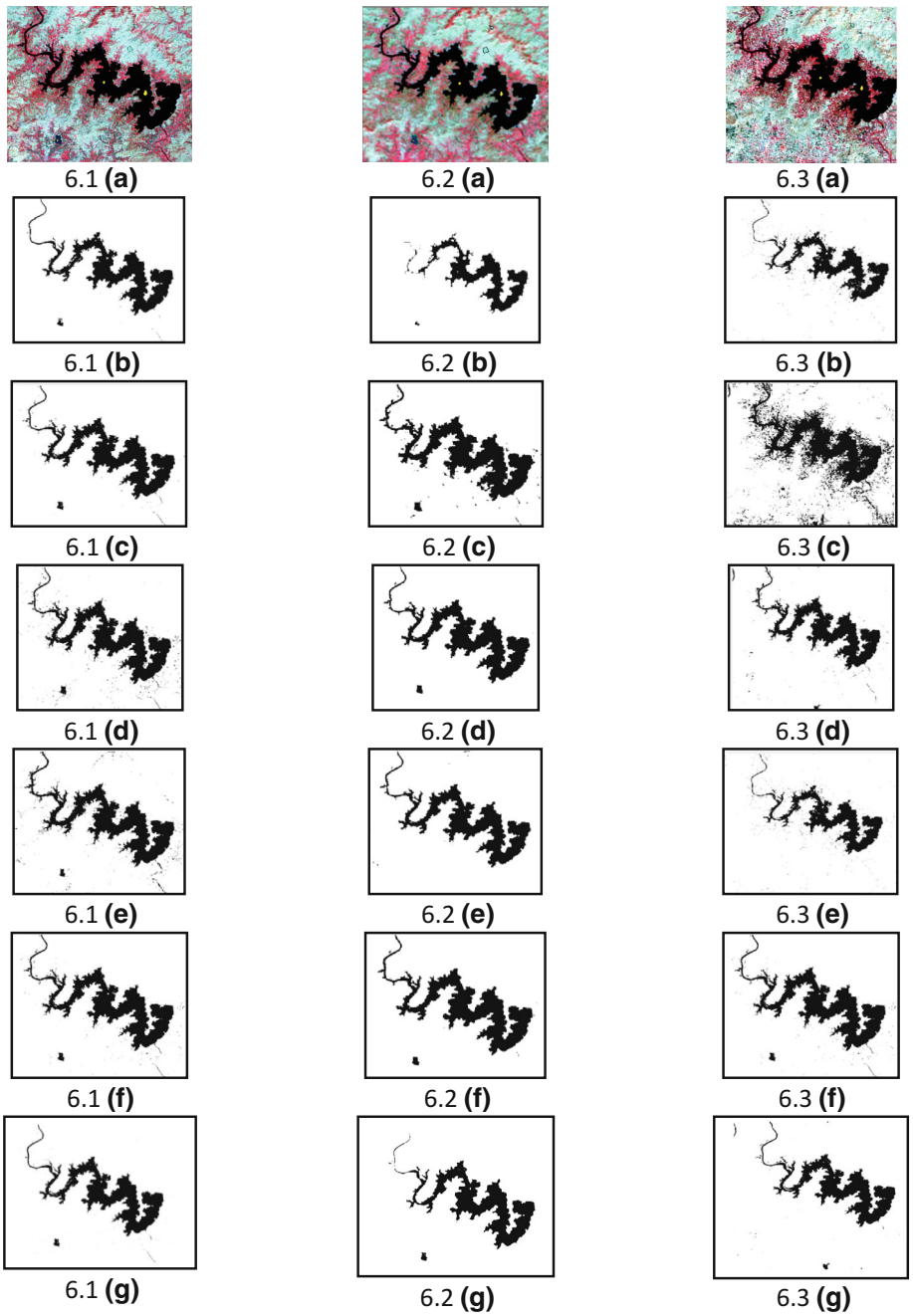

Fig. 6 Study area II: each column (*left* to *right*) represents **a** FCC of LISS-III, AWiFS and LANDSAT data (overlaid with labeled samples) along with extracted water bodies using: **b** Mahalanobis discriminant, **c** CSVC-polynomial, **d** KNN, **e** RF, **f** ANN and **g** NDWI methods, respectively

fast, the subjective threshold becomes the major constraint in applying NDWI method for water body extraction.

Conclusion

This paper uses advanced methods of supervised classification along with traditional method to extract water bodies from satellite images. The effect of classification has been considered in each case and it has been found out that the accuracy obtained is very high, even when the number of training samples provided is low (approximately 10–15 pixels for water body classification). This study implements sensitivity analysis with respect to images of three different study areas by the usage of three different satellite sensor data namely, LISS-III, AWIFS and LANDSAT 8, as well as with respect to supervised different methods. It has been found that methods discussed here are not sensitive to the variation

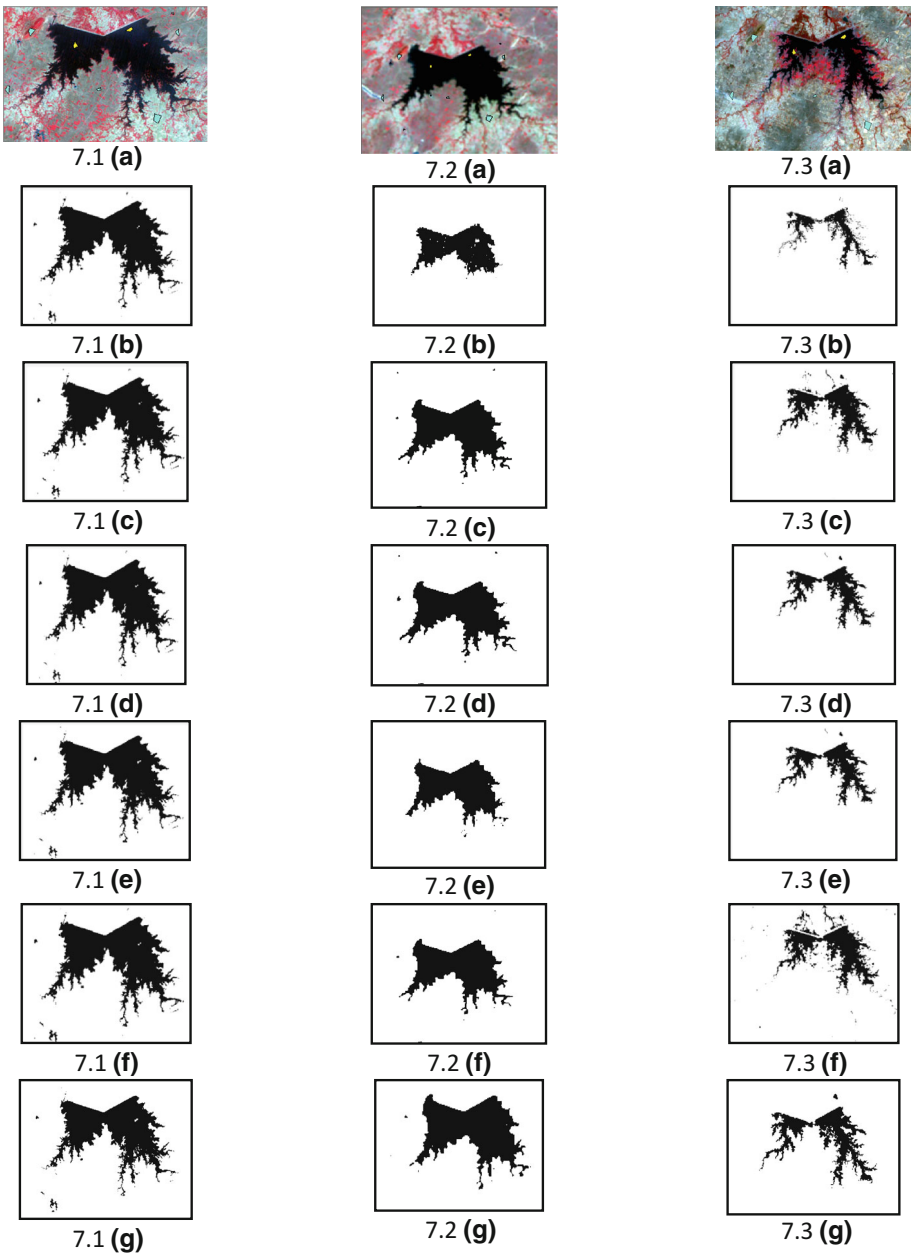

Fig. 7 Data set 3: each column (*left* to *right*) represents **a** FCC of LISS-III, AWiFS and LANDSAT data (overlaid with labeled samples) along with extracted water bodies using: **b** Mahalanobis discriminant, **c** CSVC-polynomial, **d** KNN, **e** RF, **f** ANN and **g** NDWI methods, respectively

of input satellite image for extracting water bodies. The comparative study shows the efficiency of supervised classification methods in extracting water bodies from remote sensing imagery.

Acknowledgements Authors acknowledge Dr. V. M. Chowdary, RRSC-East, ISRO for his inspiration towards the present work.

References

Althausen JD (2002) What remote sensing system should be used to collect the data? Manual of geospatial science and technology. Taylor and Francis, New York, pp 276–297

Amarsaikhan D, Douglas T (2004) Data fusion and multisource image classification. Int J Remote Sens 25:3529–3539

Atkinson PM, Tatnall ARL (1997) Neural networks in remote sensing. Int J Remote Sens 18:699–709

Barandela R, Juarez M (2002) Supervised classification of remotely sensed data with ongoing learning capability. Int J Remote Sens 23:4965–4970

Barnsley MJ (1999) Digital remote sensing data and their characteristics. In: Longley P, Goodchild M, Maguire DJ, Rhind DW (eds) Geographical information systems: principles, techniques, applications, and management, vol 2. Wiley, New York, pp 451–466

Ben-Hur A, Horn David, Siegelmann Hava, Vapnik Vladimir (2001) Support vector clustering. J Mach Learn Res 2:125–137

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Chang Chih-Chung, Lin Chih-Jen (2013) A library for support vector machines. Department of Computer Science, National Taiwan University, Taipei

Chen D, Stow DA (2002) The effect of training strategies on supervised classification at different spatial resolution. Photogramm Eng Remote Sens 68:1155–1162

Chiarandini M (2013) Gaussian discriminant analysis Naive Bayes. Introduction to Machine Learning, Department of Mathematics and Computer Science, University of Southern Denmark

Congalton RG, Green K (2009) Assessing the accuracy of remotely sensed data, 2nd edn. CRC Press, Boca Raton

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20 (3):273

Cover TM, Hart PE (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory 13(1):21–27

Dalrymple T (1960) Flood frequency analyses, manual of hydrology: Part 3 flood flow techniques. Geological survey water supply paper 1543-4

Dayan P, Sahani M, Deback G (1999) Unsupervised learning. MIT Encyclopedia of Cognitive Sciences

Du Q, Chang C (2001) A linear constrained distance-based discriminant analysis for hyperspectral image classification. Pattern Recogn 34:361–373

Duong ND (2012) Water body extraction from multi spectral image by spectral pattern analysis. Int Arch Photogramm Remote Sens Spatial Inf Sci 39(B8):181–186

Entezari-Maleki R, Rezaei A, Minaei-Bidgoli B (2009) Comparison of classification methods based on the type of attributes and sample size. J Converg Inf Technol 4(3):94–102

Erbek FS, Ozkan C, Taberner M (2004) Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. Int J Remote Sens 25:1733–1748

Farley BG, Clark WA (1954) Simulation of self-organizing systems by digital computer. IRE Trans Inf Theory 4(4):76–84

Fern´andez-Delgado M, Cernadas E, Barro S (2014) Do we need hundreds of classifiers to solve real world classification problems? J Mach Learn Res 15(1):3133–3181

Fitzgerald RW, Lees BG (1993) Assessing the classification accuracy of multisource remote sensing data. Remote Sens Environ 47 (3):362–368

Foody GM, Mathur A (2004) Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. Remote Sens Environ 93:107–117

Foody GM, McCulloch MB, Yates WB (1995) Classification of remotely sensed data by an artificial neural network: issues related to training data characteristics. Photogramm Eng Remote Sens 61:391–401

Friedl MA, Brodley CE (1997) Decision tree classification of land cover from remotely sensed data. Remote Sens Environ 61:399–409

Hastie T, Tibshirani R (1996) Discriminant analysis by gaussian mixtures. J R Stat Soc Ser B (Methodol) 58(1):155–176

Ho TK (1995) Random decision forests. In: Proceedings of the 3rd international conference on document analysis and recognition, Montreal, QC, 14–16 August 1995, pp 278–282

Ho TK (1998) The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell 20(8):832–844

Hogland J, Billor N, Anderson N (2013) Comparison of standard maximum likelihood classification and polytomous logistic regression used in remote sensing. Eur J Remote Sens 46:623–640. doi:10.5721/EuJRS20134637

Jensen JR (2000) Remote sensing of the environment: an earth resource perspective, 2nd edn. Prentice Hall, Upper Saddle River, p 544

Ji L, Zhang L, Wylie B (2009) Analysis of dynamic thresholds for the normalized difference water index. Photogramm Eng Remote Sens 75:1307–1317

Jiang H, Feng Min, Zhu Yunqiang, Ning Lu, Huang Jianxi, Xiao Tong (2014) An automated method for extracting rivers and lakes from landsat imagery. Remote Sens 6:5067–5089

Johnson RA, Wichern DW (2007) Applied multivariate statistical data analysis, 6th edn. Prentice Hall, Upper Saddle River, p 773

Kamavisdar P, Saluja S, Agrawal S (2013) A survey on image classification approaches and techniques. Int J Adv Res Comput Commun Eng 2(1):1005–1009

Keller A, Saktivadivel R, Secler D (2000) Water scarcity and the role of storage in development. Research Report 39

Liaw A, Wiener M (2002) Classification and regression by random forest. R News 2(3):18–22. ISSN: 1609-3631

Lillesand TM, Kiefer RW (1994) Remote sensing and image interpretation, 3rd edn. Wiley, New York, p 750

Lu D, Weng Q (2007) A survey of image classification methods and techniques for improving classification performance. Int J Remote Sens 28(5):823–870

McCulloch W, Pitts Walter (1943) A logical calculus of ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133

McFeeters S (1996) The use of the normalized difference water index (NDWI) in the delineation of open water features. Int J Remote Sens 17:1425–1432

Mishra K, Prasad PRC (2014) Automatic extraction of water bodies from landsat Imagery using perceptron model. J Comput Environ Sci 2015:1–8

Otsu N (1979) A threshold selection method from Gray-level histograms. IEEE Trans Syst Man Cybern 9(1):62–66

Otukei J, Blaschke T (2010) Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. Int J Appl Earth Obs Geoinf 12:S27–S31

Perumal K, Bhaskaran R (2010) Supervised classification performance of multispectral images. J Comput 2(2):124–129

Press SJ, Wilson S (1978) Choosing between logistic regression and discriminant analysis. J Am Stat Assoc 73(364):699–705. doi:10.2307/2286261

Roughgarden J, Running SW, Matson PA (1991) What does remote sensing do for ecology? Ecology 72(6):1918–1922

Rundquist DC, Lawson MP, Queen LP, Cerveny RS (1987) The relationship between summer-season rainfall events and lake-surface area. J Am Water Resour Assoc 23:493–508

Xu Han-Qiu (2005) A study on information extraction of water body with the modified normalized difference water index (MNDWI). J Remote Sens 9(5):589–595

Xu H (2006) Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int J Remote Sens 27:3025–3033

Yang H, Wang Zongmin, Hongling Z, Yu G (2011) Water body extraction methods study based on RS and GIS. Proc Environ Sci 10:2619–2624

Zhang Z, Prinet V, Song De M (2003) Water body extraction from multi-source satellite images. In: Proceedings of the Geoscience and Remote Sensing Symposium (IGARSS'03), 6:3970–3972