International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: http://www.tandfonline.com/loi/tres20

The application of ensemble techniques for land-cover classification in arid lands

Marwa Waseem A. Halmy & Paul E. Gessler

To cite this article: Marwa Waseem A. Halmy & Paul E. Gessler (2015) The application of ensemble techniques for land-cover classification in arid lands, International Journal of Remote Sensing, 36:22, 5613-5636, DOI: 10.1080/01431161.2015.1103915

To link to this article: http://dx.doi.org/10.1080/01431161.2015.1103915

Published online: 16 Nov 2015.

International Journal of Remote Sensing, 2015
Vol. 36, No. 22, 5613–5636, http://dx.doi.org/10.1080/01431161.2015.1103915

The application of ensemble techniques for land-cover classification in arid lands

Marwa Waseem A. Halmy[a]* and Paul E. Gessler[b]

[a]*Department of Environmental Sciences, Faculty of Science, Alexandria University, PO Box 21511, Alexandria, Egypt;* [b]*Department of Forest, Rangeland and Fire Sciences, University of Idaho, PO Box 441133, Moscow, ID 83844-1133, USA*

(*Received 1 May 2014; accepted 26 September 2015*)

Robust classification approaches are required for accurate classification of complex land-use/land-cover categories of desert landscapes using remotely sensed data. Machine-learning ensemble classifiers have proved to be powerful for the classification of remotely sensed data. However, they have not been evaluated for classifying land-cover categories in desert regions. In this study, the performance of two machine-learning ensemble classifiers – random forests (RF) and boosted artificial neural networks – is explored in the context of classification of land use/land cover of desert landscapes. The evaluation is based on the accuracy of classification of remotely sensed data, with and without integration of ancillary data. Landsat-5 Thematic Mapper data captured for a desert landscape in the north-western coastal desert of Egypt are used with ancillary variables derived from a digital terrain model to classify 13 different land-use/land-cover categories. Results show that the two ensemble methods produce accurate land-cover classifications, with and without integrating spectral data with ancillary data. In general, the overall accuracy exceeded 85% and the kappa coefficient (κ) attained values over 0.83. The integration of ancillary data improved the performance of the boosted artificial neural networks by approximately 5% and the random forests by 9%. The latter showed overall higher accuracy; however, boosted artificial neural networks showed better generalization ability and lower overfitting tendencies. The results reveal the merit of applying ensemble methods to integrated spectral and ancillary data of similar desert landscapes for achieving high classification accuracies.

1. Introduction

Machine learning refers to algorithms that are used in data analysis and pattern recognition by iteratively learning from training observations (DeFries and Chan 2000). In remote sensing, machine-learning methodologies, especially ensemble approaches, have gained popularity over the last decade (Miao et al. 2012). Camps-Valls (2009) provided a review of machine-learning techniques used in remote sensing and highlighted the potential of new promising learning approaches that have not been considered in remote-sensing application. The application of machine-learning approaches for image classification problems is seen as a 'competitive alternative' to conventional classification approaches (Zhang 2000).

Combining classifiers in an ensemble approach involves the iterative use of a base classifier followed by combining the results according to a weighted or unweighted voting process (Breiman 1996; Briem, Benediktsson, and Sveinsson 2002; Dietterich 2000; McInerney and Nieuwenhuis 2009). Combining the results of all the iterations provides

*Corresponding author. Email: marwa.w.halmy@alexu.edu.eg

a more stable final classification output (Cunningham, Carney, and Jacob 2000). Ensemble methods could be applied to any learning algorithm, and they can include one or more type of learning algorithm (Chan and Paelinckx 2008); however, they are typically composed of one type of learning algorithm (Miao et al. 2012). Ensemble methods are used in remote-sensing applications for improving the classification accuracy (Rokach 2005) and for overcoming the instability of some machine-learning classifiers (Cunningham, Carney, and Jacob 2000; Dietterich 2000) such as artificial neural networks (ANNs) (Canty 2009) and decision trees (DTs) (Breiman 2001). Ensembles of certain base classifiers have been found to provide more accurate results compared to the use of a single base classifier (e.g. boosted DTs *vs*. single DT) of the same type (Gislason, Benediktsson, and Sveinsson 2006; Rodriguez-Galiano et al. 2012). The performance of each ensemble approach depends on the base classifier used.

Bagging (Breiman 1996) and boosting (Freund and Schapire 1996) are the most commonly used ensemble methods (Chan and Paelinckx 2008; Dietterich 2000; Miao et al. 2012). In bagging (a.k.a. 'bootstrap aggregation') (Polikar 2006), bootstrap replicas are drawn from the original training data set with replacement. Each training data replica is then used in a classification iteration using a machine-learning algorithm (e.g. DT or ANNs). The results from all the classification iterations are then combined where classes are finally assigned based on the most common vote in all the iterations (Franklin 2009; Polikar 2006). Random forests (RF) technique (Breiman 2001) is another popular ensemble method that is considered a form of bagging (Benediktsson, Chanussot, and Fauvel 2007; Franklin 2009). As in bagging, RF trains DTs on bootstrapped samples drawn of the original training data (Breiman 2001), but unlike bagging, a random subset of the input variables is used to split nodes in each DT (Polikar 2006). The number of DTs used in a RF model and the number of variables used to split nodes of each DT in the ensemble are user-defined parameters that are selected based on the case under study. Each DT is trained on a bootstrapped sample of the original training data (Gislason, Benediktsson, and Sveinsson 2006; Rodriguez-Galiano et al. 2012).

Boosting uses iterative re-training, where the incorrectly classified observations are given more weight in the successive iteration. This way, the variance and the bias of the classification are reduced, resulting in a more accurate classification (Dietterich 2000). Boosting is considered more accurate compared to other ensemble methods; however, on the other hand, it is considerably slower and sensitive to noise (Briem, Benediktsson, and Sveinsson 2002). There are many boosting algorithms (Briem, Benediktsson, and Sveinsson 2002), the most commonly known of which is the adaptive boosting (AdaBoost) introduced by Freund and Schapire (1996). It has been mostly applied to DT (Ridgeway 1999); however, it could be applied to other learning algorithms such as ANNs (e.g. Schwenk and Bengio 2000; Canty 2009). Boosting ANNs is not common in remote-sensing applications, although it has shown significant improvement over conventional (e.g. maximum likelihood) algorithms and other machine-learning classifiers such as support vector machine when used for the classification of remotely sensed data (Canty 2009).

The 'multilayer perceptron' neural networks (MLPs) were developed by Rumelhart, Hinton, and Williams (1986) for pattern-recognition applications and have since been extensively used in many other applications. ANNs are, by far, the most widely used type of neural network in remote-sensing applications (Bischof, Schneider, and Pinz 1992; Kavzoglu and Mather 2003; Kotsiantis 2007; Mas 2004; Paola and Schowengerdt 1995; Pijanowski et al. 2002; Serpico, Bruzzone, and Roli 1996). ANNs were found to outperform conventional classifiers in land-cover classification (Bischof, Schneider, and Pinz 1992; Serpico, Bruzzone, and Roli 1996; Tong, Zhang, and Liu 2010; Yool 1998). A review of the applications of ANNs for classification of remotely sensed data, their limitations, and perspectives can be found in Mas and Flores (2008). The application of ensembles of ANNs to remotely sensed data have resulted in significant increases in land-use/land-cover (LULC) classification accuracies over the use of single neural networks (He, Kong, and Shen 2006); popular conventional methods such as maximum likelihood; and other machine-learning algorithms like support vector machine approaches (He, Kong, and Shen 2006; Canty 2009).

Integrating environmental data with spectral data often enhances the accuracy of LULC classification (Benediktsson, Chanussot, and Fauvel 2007), but this is not always possible with conventional classification methods (Briem, Benediktsson, and Sveinsson 2002; Gislason, Benediktsson, and Sveinsson 2006). As machine-learning approaches, ensemble classifiers have the advantage of being applicable to multiple data types, allowing the integration of spectral and ancillary data of different types (Horning 2010). Integration of ancillary data (e.g. land-surface parameters, soil data, etc.) with spectral data could be useful when the study area is located in a highly reflective desert ecosystem. The high reflective background and the low spectral variability of surface features in deserts make the classification of land cover based on spectral data alone challenging (Halmy 2012; Halmy et al. 2015). The application of ensemble-based classification approaches such as bagging, boosting, and RF is a valuable alternative in such instances. However, ensemble methods that have been applied and tested in different types of landscapes and ecosystems have been in mostly non-desert areas including temperate and humid ecosystems (Halmy 2012). This may be due to challenges with field work and, perhaps, lack of funding for studies in desert or arid areas.

For example most of the studies that employed RF for LULC classification are in temperate (Castaings et al. 2010; Ismail and Mutanga 2011; McInerney and Nieuwenhuis 2009; Pal 2005; Rodriguez-Galiano et al. 2012; Sluiter and Pebesma 2010; Waske et al. 2009) or tropical areas (Sesnie et al. 2010). In the context of these studies, RF was compared to conventional classifiers (McInerney and Nieuwenhuis 2009; Sluiter and Pebesma 2010) and to machine-learning algorithms (Ismail and Mutanga 2011; Miao et al. 2012; Pal 2005, 2006; Rodriguez-Galiano et al. 2012; Sesnie et al. 2010), and was found to outperform both conventional and other machine-learning approaches. Few studies have compared the boosted ANNs to other classifiers. Canty (2009) compared the boosted ANNs to maximum likelihood, support vector machine (SVMs), and boosted trees. However, we are unaware of studies that have compared boosted ANNs to RF, particularly in the context of classifying land cover of desert and semi-desert landscapes. The current study assesses and compares the accuracy of these two ensemble classifiers for the classification of LULC of a desert landscape in Egypt, with and without integration of ancillary data.

2. Materials and methods

2.1. Study area

The study area lies within the jurisdiction of the Matrouh governorate that occupies the northern part of the Western Desert of Egypt (Figure 1). This study focuses on approximately 2800 km^2 of the north-western coastal desert that is located between 30° 10′–30° 55′ N and 28° 55′ to 29° 25′ E. The Egyptian north-western coastal desert area is generally a topographically featureless plain that is cut in some places by depressions and oases. The geological formations of this part of the Western Desert are mainly Quaternary and Tertiary in age. Two Miocene rock units can be distinguished in the area – the

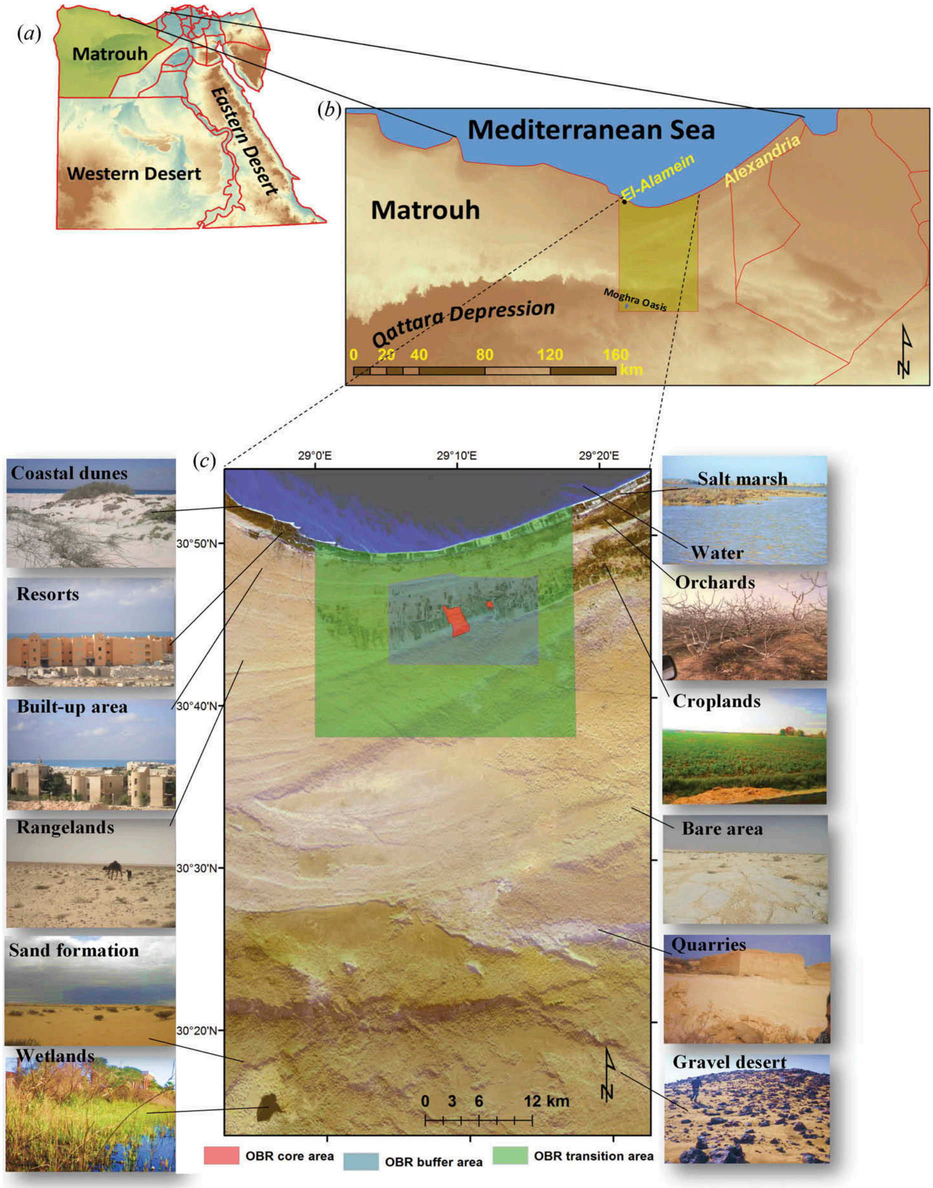

Figure 1. Location of study area: (*a*) the administrative boundaries of the Egyptian governorates; (*b*) part of the north-western coastal desert showing location of the study area; and (*c*) subset of Landsat TM imagery form 2011 representing the study area and showing locations for the studied LULC classes and the different zone of Omayed biosphere reserve (OBR).

Marmarica limestone formation and the Moghra clastic formation. The Marmarica formation that forms the surface of the coastal area is composed of chalky white limestone as an upper member and grey calcarenites as a lower member. The Moghra formation forms the surface of the southern area that encompasses the Moghra oasis. This formation is a clastic unit that has varying sand:shale ratio typically of approximately 3.5:1, and it outcrops the surface at the southern edge of the inland plateau.

The Moghra Oasis is bounded to the north by an escarpment that leads to gravelly slopes formed as a result of water erosion and to the west by sand dunes and sand formations (De Cosson 1935; Said 1962). Studies of the area (e.g. Ayyad and Le Floc'H 1983; Said 1962; Salem 1989, 2003a, 2003b, 2005; ROSELT/OSS 2005) have distinguished a number of geomorphological units in an order from north to south; northern coastal dunes that are aligned parallel to the coast and composed of oolitic limestone (Ball 1939), followed by a wide northern coastal plain, then an intermediate tableland (the inland plateau), gravelly slopes, and sandy-gravelly plain that surrounds Moghra Oasis. A series of elongated calcareous ridges that are aligned in a northeast–southwest direction occur in the northern coastal plain alternating with sandy depressions. The inland plateau is characterized by hard calcareous limestone that is covered in some places by a thin layer of sandy or loamy deposits. The plateau slopes southward to an undulated gravelly plain formed by deposits of mixed aeolian and colluvial origin (Ayyad and Ghabbour 1993; Ayyad and Le Floc'H 1983; ROSELT/OSS 2005). The area has an arid climate with mean annual precipitation of 100–150 mm year^{-1}. The temperature and precipitation decrease in a north–south trend (Abdel Razik 2008).

2.2. LULC categories

The use of a standard LULC classification scheme helps facilitate future comparative studies. The LULC classification system used here is an adaptation of Anderson et al.'s (1976) system. Anderson's scheme was developed by the United States Geological Survey (USGS) to serve as a framework for the classification of remotely sensed data for the entire globe. Level II of Anderson's scheme suits the spatial resolution of Landsat TM (Ridd 1995). Based on level II of Anderson, 13 LULC categories were distinguished in the study area (Table 1 and Figure 1).

Table 1. Land use/Land cover classes recognized in the study area in 2011 defined according to Anderson et al. (1976) level II with adaptation to suit the study area.

Anderson et al. (1976) LULC classes	Anderson et al. (1976) Code	Current study LULC classes	Current study code	Colour code
Residential	**11**	Urban & built-up land	**BU**	
Other urban & built-up land	**17**	Resorts	**RE**	
Croplands	**21**	Croplands	**CR**	
Orchards & other agricultural lands	**22**	Orchards & other agricultural lands	**OA**	
Rangeland	**3**	Rangeland	**RL**	
Water	**5**	Water	**WT**	
Non-forested Wetlands	**61**	Reed swamp	**RS**	
		Salt marshes	**SM**	
Beaches	**72**	Coastal dunes	**CD**	
Sandy area other than beaches	**73**	Sand formation	**SF**	
Bare exposed area	**74**	Bare exposed area	**BA**	
Quarries	**75**	Quarries	**QR**	
Gravel Desert	**75**	Gravel desert	**GD**	

2.3. Spectral and ancillary data

Because of the similarity in the spectral response among many desert land features, it is hard to separate land-cover classes using spectral data alone (He, Kong, and Shen 2006). Although the study area has various LULC categories, their classification based on spectral data alone is often challenging due to spectral similarities. As a part of the coastal desert of Egypt, most of the land-cover categories in the area are spectrally heterogeneous and have overlapping spectral properties. The exposed bare rock areas and the quarries are spectrally similar as both are composed of limestone rock. These challenges make it difficult to achieve accurate classifications using spectral data alone; therefore, additional ancillary data will be used to supplement the Landsat spectral data to aid in separating these land-cover categories (Weng 2002).

A Landsat TM 5 scene (path/row 178/39) was acquired on 5 March 2011 and provided the spectral data for the classification of the LULC categories in the area. The equations and the rescaling factors provided by Chander, Markham, and Helder (2009) were applied to convert the Landsat-calibrated DNs to absolute units of at-sensor spectral radiance and, then, to top-of-atmosphere reflectance.

Many studies indicate that the inclusion of texture data improves classification accuracy (Franklin et al. 2000). Therefore, texture variables were included in the classification process to aid detection of built-up and topographic features (Gong, Marceau, and Howarth 1992; Kiema 2002; Lu, Hetrick, and Moran 2010). A kernel of size 3 × 3 pixels was used to derive texture measures from the near-infrared (NIR) band based on the grey level co-occurrence matrix method (GLCM) of Haralick, Shanmugam, and Dinstein (1973). The NIR band was used because it exhibits better contrast between the different LULC classes in the study area than the other bands. The textural measures derived included six second-order textural features (homogeneity, contrast, dissimilarity, entropy, second moment, and correlation) and two first-order textural features (mean and variance). The textural features were computed for four directions, 0°, 45°, 90°, and 135°, and the average was used in the analysis. Furthermore, the tasseled cap spectral indices of greenness, brightness, and wetness (Kauth and Thomas 1976) were derived from the Landsat data and included in the analysis.

Approaches for integrating ancillary data with spectral data have been found to improve LULC classification accuracy (He, Kong, and Shen 2006; Sesnie et al. 2008; Stuckens, Coppin, and Bauer 2000). The ancillary variables used in the study to help in differentiating between LULC included topographic variables derived from a 90 m grid-spacing digital elevation model (DEM) and distance to the sea. The proximity to the sea was included here because field work revealed a north–south trending pattern of LULC classes in the area. The land-surface parameters included in the study (Table 2) were calculated from the Shuttle Radar Topographic Mission (SRTM) DEM data V4.1 (Jarvis et al. 2008). The DEM data were used to derive other land-surface parameters, such as slope, a terrain roughness index (TRI) that provides an indication about the level of undulation and the complexity of the surface (Olaya 2009), topographic wetness index (TWI) that indicates the water accumulation possibility of a pixel (Moore et al. 1993; Gessler et al. 1995, 2000), and a slope length and steepness (LS) factor. All the land-surface parameters were derived using the Automated System Geoscientific Analyses version 2.0.7 (SAGA 2011).

2.4. Classification

Two ensemble classifiers – RF and boosted ANNs – are compared here in the context of mapping LULC in desert areas. The comparison involves assessing the difference in their

Table 2. Spectral and ancillary data included in the classification.

Variable type	Variable	Abbreviation
Spectral	Spectral bands 1–5 & 7	B1–B6
	Texture	Tx1–Tx8
	NDVI	NDVI
	Simple ratio (IR:R)	IR.R
	Tasselled cap brightness	Brightness
	Tasselled cap greenness	Greenness
	Tasselled cap wetness	Wetness
Ancillary	Proximity to the sea	ProxtoSea
	Elevation	Elv
	LS factor	LS
	Slope	Slope
	Topographic roughness index (TRI)	TRI
	Topographic wetness index (TWI)	TWI

performance with and without ancillary data (Table 2). Classifying LULC in the area was implemented using three sets of data, referred to here as data domains. The first set included spectral data of six Landsat satellite image bands (bands 1–5 and 7) alone; the thermal band was excluded because it has coarser resolution compared to the rest of the bands. The second included spectral data integrated with ancillary data derived from the DEM, and the third included 10 principal component (PCs) bands resulting from applying principal component analysis (PCA) to the whole set of variables representing spectral and ancillary data (Table 2). The greater the number of variables used in the classification, the greater the time required for processing the data by the classifiers. PCA was used for reduction of the dimensionality of the set of data employed in the classification. These first 10 PCs were chosen because they accounted for more than 95% of the variation in the whole set of the integrated data. The boosted ANNs algorithm of Canty (2009) was used to perform classification using boosted ANNs and was implemented within the framework of ENVI 4.8 (Exelis Visual Information Solutions 2010). Classification using random forests was conducted using the 'randomForest' package (Liaw and Wiener 2002) within the statistical software R 2.13.1 (R Core Development Team 2011).

For each data domain, trials were made to reach the optimum structure of the two ensemble classifiers.

2.4.1. RF models Optimization

RF is a 'tree-based' classifier that requires setting only two user-defined parameters when performing classification: 1) number of trees to be grown in the RF model and 2) the number of variables to be used at split nodes (Pal 2005; Rodriguez-Galiano et al. 2012). There is no predefined rule for deciding on the optimal number of trees to be used. A small number of trees may result in low classification accuracies. On other hand, increasing the number of trees beyond a certain limit (depending on the case at hand) might not reduce the classification error, but will increase the time spent in the classification process (Rodriguez-Galiano et al. 2012). In the current study, different numbers of trees were tried (50, 250, 500, 750, 1000, and 1250) to arrive at the optimum model size.

The variables used for splitting nodes at each tree are selected at random with the intent of minimizing the correlation between the trees in the RF model (Gislason, Benediktsson, and Sveinsson 2006; Horning 2010; Rodriguez-Galiano et al. 2012). The fewer the variables used to split nodes, the fewer computations the algorithm performs and the lower the correlation between trees in the RF model. But using few variables to split nodes could result in weak models with lower predictive power (Gislason, Benediktsson, and Sveinsson 2006; Horning 2010). Therefore, the choice of the number of variables to be used at split nodes should be chosen carefully. Some studies decide on the number of variables m according to the rule $m = \sqrt{k}$, where k is the total number of variables used in the classification (Gislason, Benediktsson, and Sveinsson 2006; Chan and Paelinckx 2008; Waske et al. 2009). For the current study, different numbers of split variables m were tried for each data domain using different RF models of different size. For the spectral and PCs data domains, $m = 1$ up to 6 were experimented, whereas, for the multisource data domain, $m = 2$ up to 10 were experimented. Based on the optimization trials, RF models with 500 trees and 2 variables at split nodes were applied to the spectral and PCs data domains; for the multisource data, domain RF models of size 500 trees and four split variables were applied.

2.4.2. Boosted ANNs models Optimization

The generalization power of ANNs and, thus, the resulting classification accuracy are contingent on factors that include training data size, network training time, and network design (Pal and Mather 2003). The network design or architecture is determined by the number of hidden layers and the number of nodes in each hidden layer (Kavzoglu and Mather 1999). Reaching the optimal design of ANNs could be an elusive task (Moody 1991). Because ANNs architecture influences the classification accuracy, it should be set up carefully (Maier and Dandy 2000; Pal and Mather 2003). Most applications that involve classification of remotely sensed data have used a three-layered fully interconnected feed-forward MLP network with a single hidden neuron layer (Paola and Schowengerdt 1995). The ANNs used here is a four-layer feed-forward neural network based on the multilayer perceptron (MLP) structure. Using two hidden layers, the number of hidden nodes per layer can be reduced compared to that required when using a single layer (Kavzoglu and Mather 2003; Aitkenhead and Aalders 2008). The network consists of an input layer, an output layer and two hidden layers and is trained by a Kalman filter algorithm, as outlined by Canty (2009). The input layer is composed of the number of variables introduced at the classification (i.e. six neurons for the six Landsat spectral bands used here). The output layer is composed of 13 neurons, corresponding to the number of the LULC classes being classified.

The number of hidden neurons to be included in each layer is a user-defined parameter, and careful selection is important (Bischof, Schneider, and Pinz 1992). Underestimating this number may lead to poor classification, while overestimating may lead to overfitting (Kotsiantis 2007; Canty 2010). There are few heuristic rules for the selection of the optimal number of nodes, most of the studies relied on trial-and-error for determining the number of neurons to be used (e.g. Paola and Schowengerdt 1995; He, Kong, and Shen 2006; Aitkenhead and Aalders 2008; Canty 2010). The decision on the optimal structure of the network and the values of the parameters involved are dependent on the nature of data being classified. The use of two hidden layers entails the use of a lower number of neurons per layer compared to the use of one hidden layer alone (Aitkenhead and Aalders 2008). The rules defined in some studies with regard to the number of hidden neurons were developed for networks that contain only one hidden layer; this might imply that applying these rules for networks with more than one hidden layer might not be appropriate.

To optimize the network architecture by selecting the optimum number of neurons in each hidden layer and the number of training epochs, many trials were undertaken to evaluate the impact of changing each parameter on the classification accuracy. Different numbers of hidden neurons were tested, taking into account some of the rules mentioned by Kavzoglu and Mather (2003) as a guideline. The two hidden layers of the boosted ANNs models were initiated with five hidden neurons in each layer, and then the number was increased with an increment of five hidden neurons up to 30 hidden neurons, while fixing the number of training cycle (epochs) at 10.

Boosting is implemented, in the current study, using the AdaBoost.M1 algorithm (Freund and Schapire 1996). For a set of training instances of size m with label $y_i \epsilon Y = \{1, 2 \ldots. k\}$, where k, in this case, is the number of land-cover classes, an initial distribution $D_1(i) = 1/m$ is used for the ith instance. The learning classifier – in this case the ANN – commences the training process using the training instances with the initial distribution D_1. The misclassification error e_t of each classification iteration $t = 1, 2 \ldots T$, will be calculated and used for updating the training instances distribution $D_t(i)$ by factor $\beta_t = e_t/(1-e)$. The new distribution $D_{t+1}(i)$ will account for the error of classification in the preceding training iteration t, where the misclassified pixels will be emphasized in the new training iteration. Thus, Adaboost directs the used classifiers in the ensemble to focus more on the difficult-to-classify instances.

The final vote for each instance resulting from the ensemble that is composed of T iterations will be weighted by a factor β_t. This factor is a function of the classification error e_t attained at each iteration t. The error-based weight, $\ln 1/\beta_t$, will be calculated for each iteration t and used to adjust class assignment for each instance in the ensemble. More details on how the Adaboost.M1 algorithm works can be found in (Freund and Schapire 1996). Training will stop if the misclassification error e_t from the training cycle t is bigger than 0.5 or by reaching a predefined number of training cycles T (Canty 2009). Different numbers of epochs (training cycles) were tested (5, 10, 20, 40, 50, and 60), while fixing the number of the neurons at 15 in each hidden layer to detect how the number of training cycles (epochs) would affect the accuracy of the classification by using boosted ANNs of a fixed number of hidden neurons.

Based on optimization trials, the boosted ANNs ensembles applied were composed of networks with two hidden layers. The number of hidden neurons used in each hidden layer varied. For the spectral data domain, networks of 15 hidden neurons per hidden layer were applied. Networks composed of 25 neurons per hidden layer were used in the ensemble applied to all the integrated spectral and ancillary data, whereas networks with 20 neurons were used for the PCs data domain.

Using the optimum structure reached, the two ensemble methods were applied to three training data sets of the same size drawn at random from the original data set. The mean of results from the three training sets were then used to compare the performance of the two ensemble methods using the different data domains.

2.5. Accuracy assessment and classifier evaluation

The ground reference data collected during field visits to the area were selected to represent the different LULC categories in the study area based on random sampling. Figure 1 shows photographs for the 13 land-cover classes used for classification. Field observations were collected in March 2011 and used to generate a reference data set.

Approximately 39 regions of interest (ROIs) representing the 13 LULC classes were identified in the Landsat scene. The ROIs included 9750 pixels, which were divided into two subsets. The first subset included approximately 60% of the data and was used for classifier training, while the second subset included the rest of the data and was used for testing and accuracy assessment. This might be expected to increase classification accuracy. However, the use of related samples is common in studies that attempt to evaluate the relative accuracy of different classification approaches (Foody 2004). Cross-validation procedures randomly split reference data to ensure that the training data are independent of the test data (Friedl et al. 2002). Cross-validation methods were used to avoid biased results and ensure spatial separation of training and testing data. The same training and testing data sets were employed for training and testing each of the two classifiers. In this study, confusion matrices were used to derive measures of classification accuracy that included overall accuracy, producer's accuracy, user's accuracy and Cohen's kappa coefficient (κ). The Kappa coefficient (a.k.a. index of agreement) accounts for the difference between the real observed agreement and that potentially due to chance (Congalton 1991; Congalton and Green 2009). To separate the classification error to quantification error and location error, the quantity and allocation disagreement (Pontius and Millones 2011) were calculated.

To statistically test the significance of differences between the maps produced by the two different ensemble methods over the different data domains, the McNemar test (Dietterich 1998) with a 95% confidence interval was conducted.

3. Results and discussion

Studies that have compared machine-learning classification algorithms, specifically RF and boosted ANNs, in the context of mapping LULC classes in desert areas are rare. Applying the two ensemble methods to the three different data domains resulted in different classification results, as can be noticed from the LULC maps produced for the study area (Figure 2). The use of spectral data alone by both classifiers resulted in overestimation of areas classified as built-up area (BU), sand formation (SF), quarries (QR) and other agriculture (OA), compared to the use of integrated spectral and ancillary data. This was more prominent in the map produced by applying RF to spectral data alone. The integration of ancillary data helped reduce the overlap between the BU and bare area (BA), SF and BA, QR and BA, and the OA class and salt marsh (SM) class.

The overall performance of the two classifiers using the spectral data alone was almost the same, based on the results of accuracy measurements (Figures 3 and 4). However, the difference in accuracy between the two classifiers, as assessed by McNemar test (Tables 3), indicates that the null hypothesis of equivalently accurate classifiers using the spectral data can be rejected at the 5% significance level (p-value of 0.05). The results indicate that the hypothesis that the RF classifier is equivalent to boosted ANNs can be rejected in favour of RF at the 5% significance level.

Furthermore, the per-class performance of the two ensemble classifiers showed slight differences using the spectral data alone compared to that after integration of ancillary data (Figures 5 and 6). The overall mean producer's and user's accuracy for all classes was almost the same for both classifiers (87.4% mean class producer's accuracy and 86.7% and 86.6% mean class user's accuracy for boosted ANNs and RF, respectively).

Using the integrated spectral and ancillary data, the performance of the two ensemble methods differed as revealed from the results of the measures of classification accuracy (Figures 3 and 4). Inclusion of the ancillary data either before or after reduction by PCA

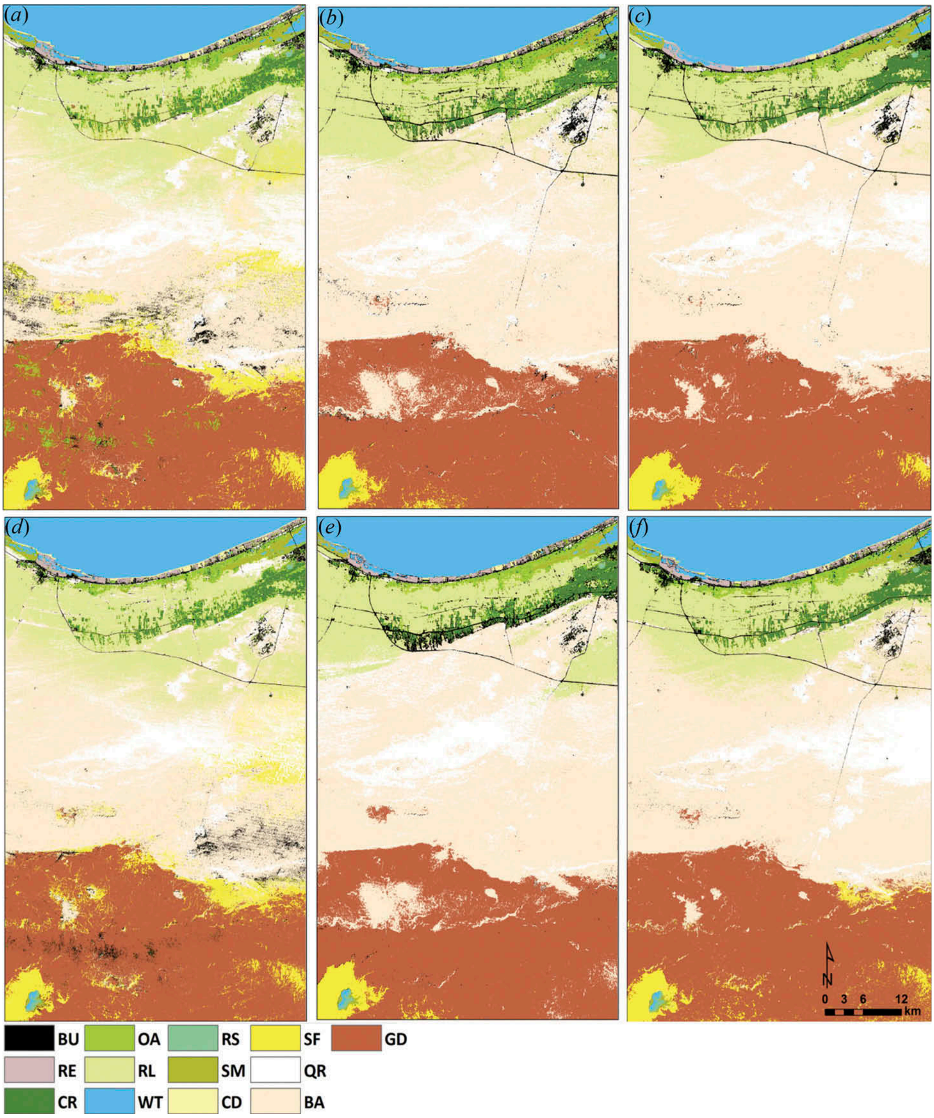

Figure 2. Land cover maps for the study area produced by the two ensemble methods, RF (top) and boosted ANNs (bottom) using spectral data (*a*) and (*d*); first 10 PCs of PCA applied to the multisource data (*b*) and (*e*); and multisource data (*c*) and (*f*). See Table 1 for LULC abbreviations.

resulted in statistically significant accuracy between the two ensemble methods (Table 3). The addition of ancillary data improved the performance of the two classifiers, resulting in higher κ and mean class accuracy values and lower classification error (Figure 3). It also reduced the quantity and allocation disagreement (Figure 4). The use of multisource data that include different forms of spatial data in addition to remotely sensed data is recommended for more accurate LULC classification applications (e.g. Briem, Benediktsson, and Sveinsson 2002; Gislason, Benediktsson, and Sveinsson 2006).

The improvement in classification accuracy as a result of inclusion of ancillary data was higher when using the RF compared to that using boosted ANNs. The integration of

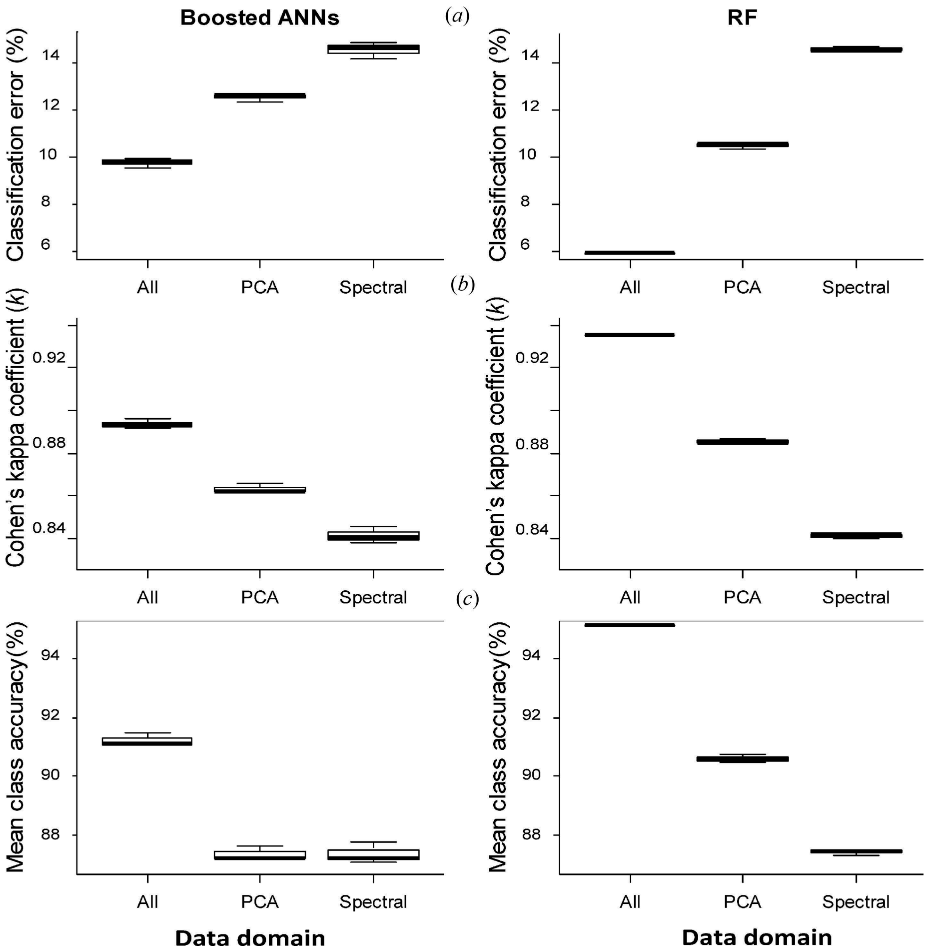

Figure 3. Box plots providing insight about the distribution of the accuracy measures of boosted ANNs and RF models using three data domains as measured by: (*a*) overall classification error; (*b*) Cohen's kappa coefficient (κ); and (*c*) mean class accuracy. Each box shows the interquartile range that contains values between 25th and 75th percentile. The middle line inside each box represents the median. The two whiskers above and below represent the maximum and the minimum values respectively.

ancillary data improved the performance of the boosted ANNs by approximately 5% while it improved the performance of RF by 9% (Figure 3). This could also be observed from the larger range in accuracy measures attained by RF over all the data domains, compared to that attained by the boosted ANNs (Figure 7). Breiman (2001) found that the error rates obtained using the RF classifier are comparable to that of boosting algorithms when it is applied to the same base classifier (DTs), with an advantage of being more robust to noise. In the current study, RF outperforms boosting using different base classifiers (ANNs in this case). The RF approach has been found to give better results than other ensemble methods such as bagging and boosting classifiers when used with multisource data. For example, the study by Benediktsson, Chanussot, and Fauvel (2007) compared the application of bagging, boosting, support vector machine, and RF in LULC classification, where RF was found to outperform the other methods. In addition, RF is

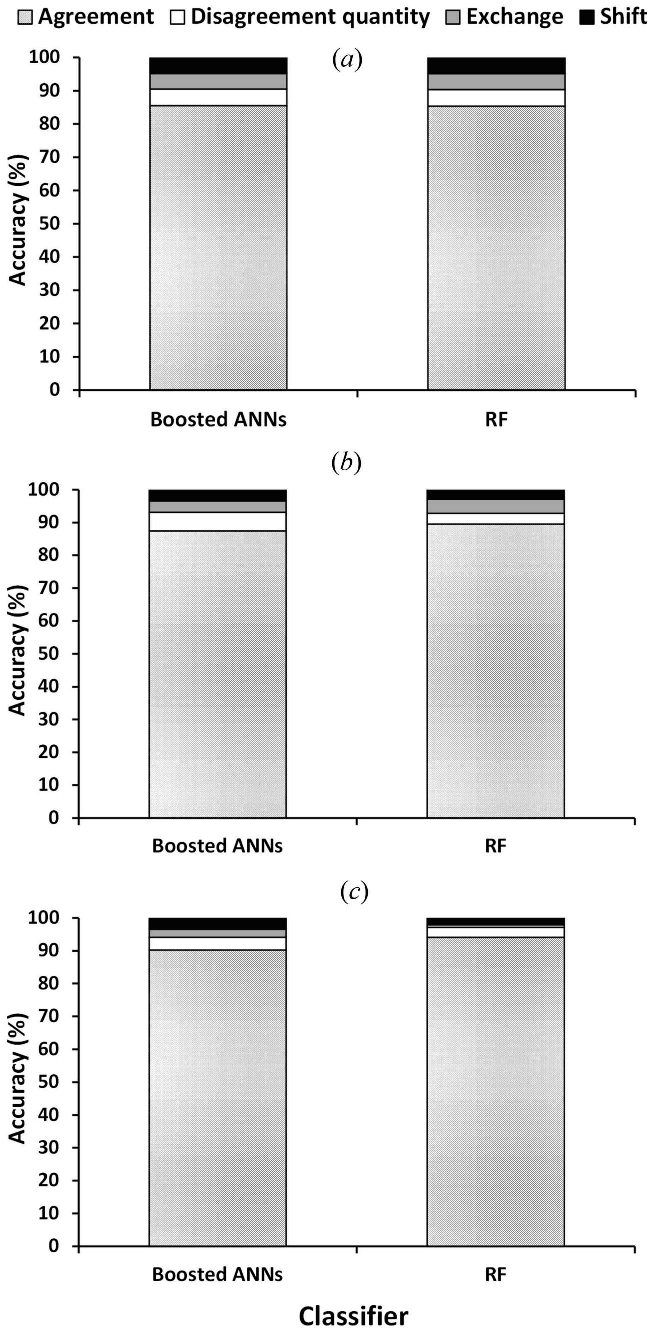

Figure 4. Classification overall agreement and disagreement quantity and disagreement allocation of the two ensemble methods over three data domains: (*a*) spectral; (*b*) first 10 PCs of PCA applied to multisource data; and (*c*) multisource data. Disagreement allocation is expressed as exchange and shift.

Table 3. McNemar's test statistic for pairwise comparison of difference in accuracy between random forests and boosted ANNs classifiers conducted over three test sets for the three data domains.

	Test set		
Data domain	Set 1	Set 2	Set 3
Spectral	46.6***	54.0***	56.0***
PCA	20.9***	24.7***	4.0*
All	54.6***	65.1***	67.3***

Note: * = p-value 0.05–0.01, ** = p-value 0.01–0.001, *** = p-value < 0.001.

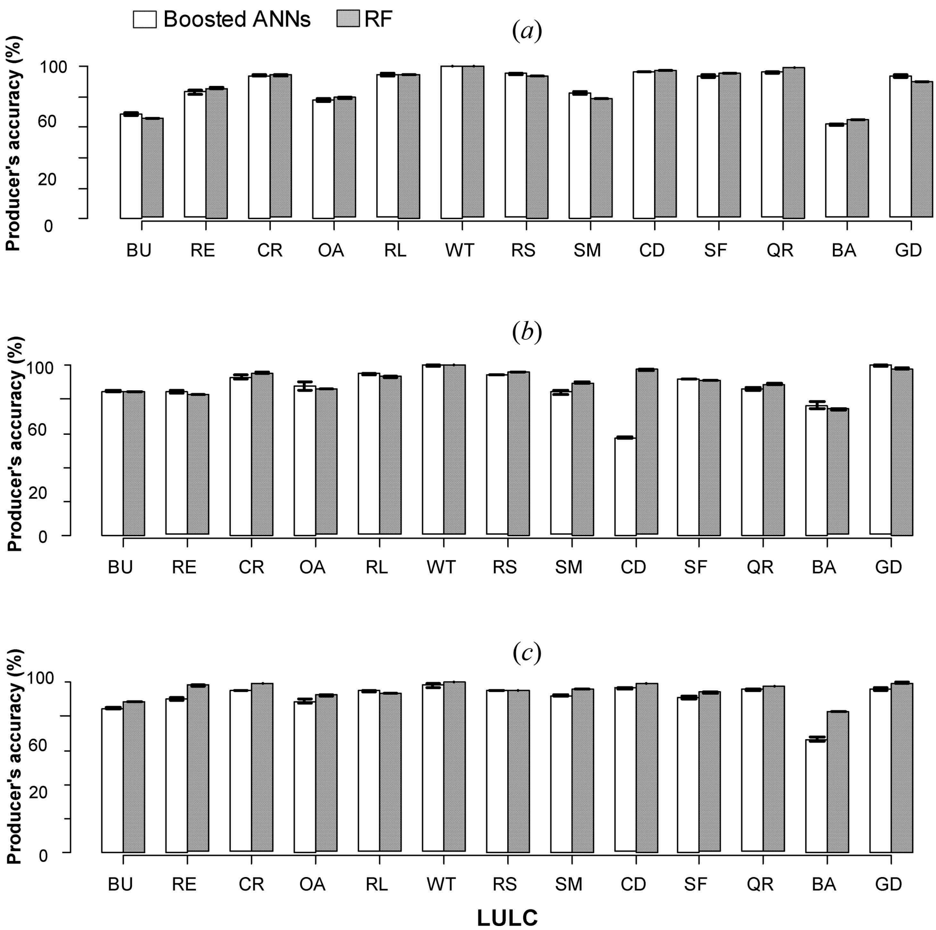

Figure 5. Comparison of the average producer's accuracy of the different LULC classes resulting by application of the two ensemble methods to three data domains: (a) spectral; (b) first 10 PCs of PCA applied to multisource data; and (c) multisource data. The error bars on the top of the histogram bars represent standard error. See Table 1 for LULC abbreviations.

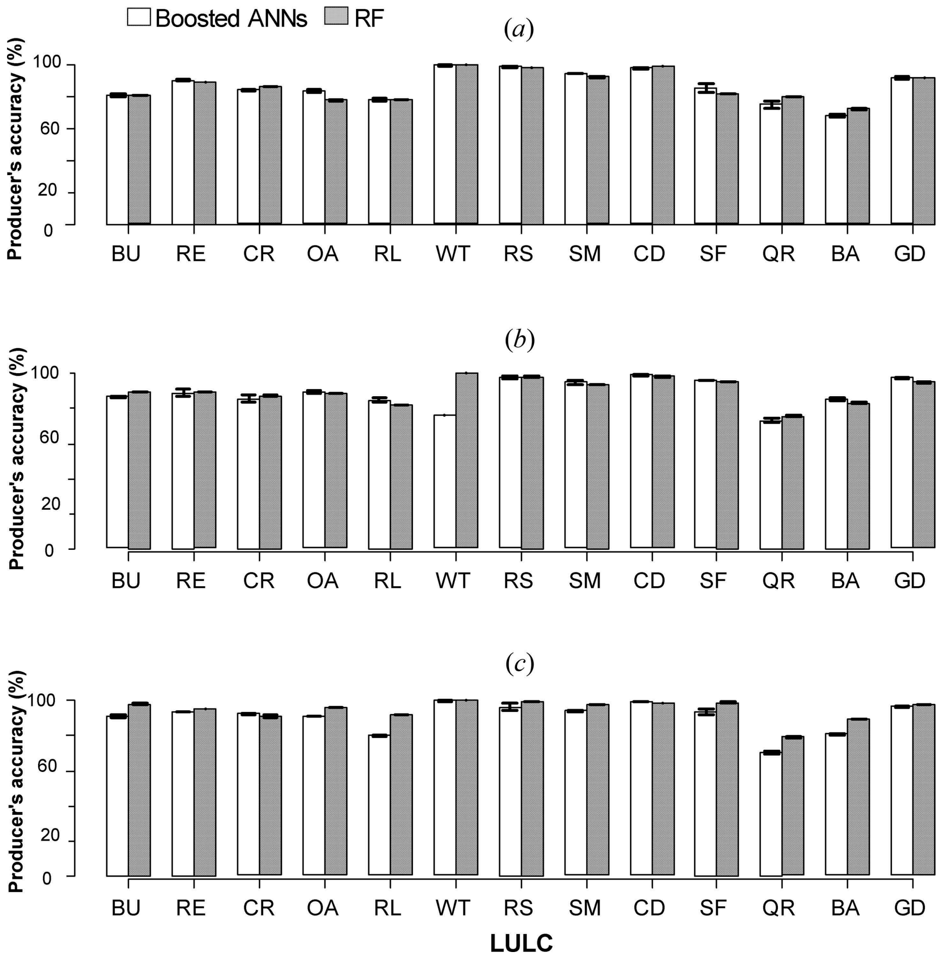

Figure 6. Comparison of the average user's accuracy of the different LULC classes resulting by application of the two ensemble methods to three data domains: (*a*) spectral; (*b*) first 10 PCs of PCA applied to multisource data; and (*c*) multisource data. The error bars on the top of the histogram bars represent standard error. See Table 1 for LULC abbreviations.

said to have better ability for deriving information from high-dimension collinear data compared to other techniques (Sluiter and Pebesma 2010).

One of the major issues facing classification of LULC using moderate resolution remotely sensed data such as Landsat data is the spectral confusion between land-cover classes that have a similar spectral response (Mas 2004). The current study area is part of a desert ecosystem where some of the land-cover classes are heterogeneous and have similar spectral properties. For example, the OA class (including orchards and other agricultural lands) has fig and olive orchards that are fenced with limestone bricks. In most cases, these orchards include houses that are built from limestone bricks. This might explain the overlap between this land-cover class and each of the BU areas and the cropland areas (CR). The rangelands (RL) in the area are sparsely covered by shrubs and dwarf shrubs that are typically overgrazed. This may lead to overlap between this class and the BA, which also contain sparse vegetation.

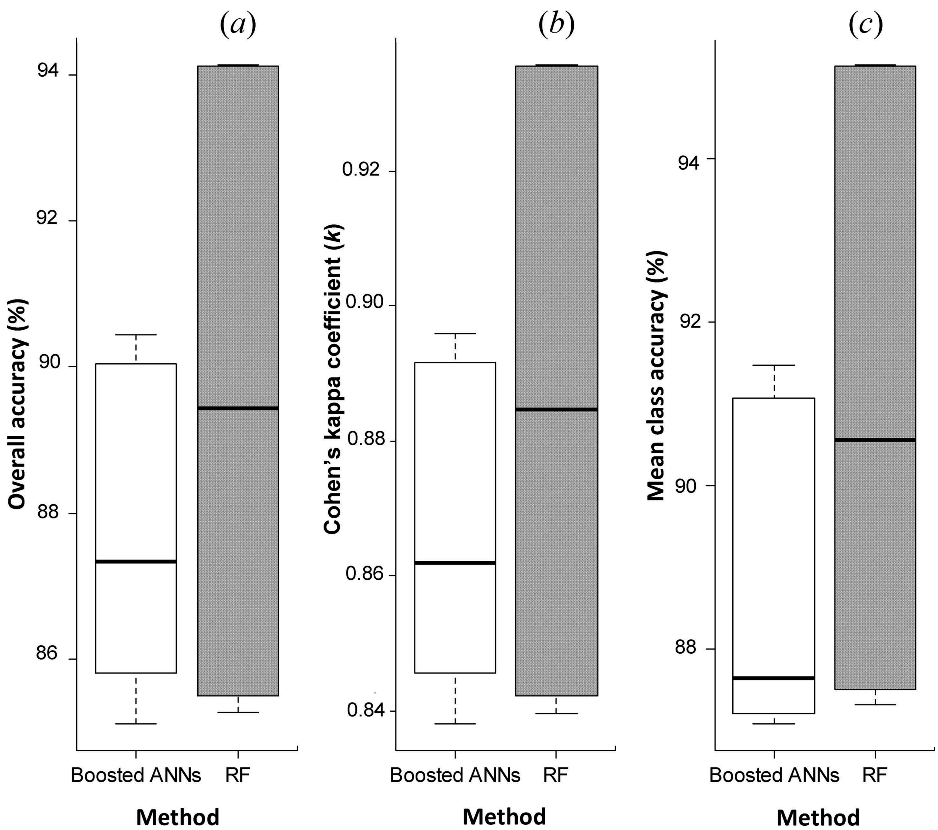

Figure 7. Accuracy of boosted ANNs and RF models over all the data domains as measured by: (*a*) overall accuracy; (*b*) Cohen's kappa coefficient (κ); and (*c*) mean class accuracy.

Integration of ancillary data with spectral data resulted in better classification of land-cover classes on a class-by-class case. Employing spectral data alone, all the classes attained producer's accuracy values higher than 80% when using the two classifiers except for BU, OA, SM, and BA (Figures 5 and 6). Homogeneous classes (e.g. WT, RS, SM, and GD) attained higher producer's classification accuracy by using boosted ANNs compared to RF. The bare area land-cover class (BA) attained the lowest producer's and user's accuracy of all the classes when using the two classifiers with high rates of both commission and omission errors.

Using all the multisource data, the producer's accuracy was improved for all the classes using the two classifiers except for BA, which attained a low producer's accuracy of 66.5% using boosted ANNs. Including ancillary data generally increased the user's accuracy for all the classes. Only the QR class attained user's accuracy lower than 80% using the two classifiers (70.6% and 79.3% for boosted ANNs and RF, respectively).

Applying the PCA to the multisource data resulted in improvement of the producer's accuracy of all the classes using the two classifiers, but led to lower producer's accuracy for the CD class using boosted ANNs. Moreover, the user's accuracy of all classes was improved, except for WT and QR, by using boosted ANNs and only QR by using RF (Figure 6).

Inclusion of ancillary data with spectral data might require the use a feature selection measure to allow choice of the best set of variables to be included in the classification process. As increasing the number of predictors may increase the time used in processing the data, it is recommended that reduction of data dimensionality be implemented. RF has the advantage over the boosted ANNs because it can be used as a feature selection tool (e.g. Sesnie et al. 2008; Chan and Paelinckx 2008) as it can provide an estimate of the variable importance using Gini Index and the oob mean decrease in accuracy as attribute selection measures. These variable importance measures (Figure 8) showed that the spectral data of the Tasseled cap bands of greenness and wetness, Landsat band 1, in addition to texture layers representing the mean and the variance of the NIR band have the strongest influence on LULC class separability. This suggests the importance of inclusion of texture data for the classification of LULC in similar areas. Moreover, the ancillary data repressing the proximity to the sea, elevation, and topographic wetness index have demonstrated high importance in the classification of the LULC classes in the study area.

Difference between test accuracies and training accuracies measures the classifier's ability to generalize. Closer values of training and test accuracies provide a good indication of classifier generality. When test accuracies deviate largely from training accuracies this indicating a high degree of overfitting of the classifier (Rogan et al. 2008). Overfitting occurs when a classifier fits the known data (training data) too closely (i.e. training data) producing high accuracy, while being less accurate in predicting new data (i.e. test data). Since the test sets and the training sets were drawn from the general pool of reference data, the differences in the test and the training accuracies indicate overfitting tendencies by both classifiers. The difference between the training and the test error attained by RF models was higher than that of the boosted ANNs over the three data domains (Figure 9). This indicates that boosted ANNs exhibit better model generalization and lower overfitting compared to RF.

Application of the two classifiers using ancillary data reduced the difference between the training and the test error, which means that addition of ancillary data improved their generalization. Although ANNs are said to suffer from overfitting, the results from this study revealed that boosting ANNs showed lower tendency for overfitting compared to RF, even though other studies (e.g. Rodriguez-Galiano et al. 2012) have indicated that RF does not suffer from overfitting.

It is worth noting that RF was faster to train compared to the boosted ANNs that carried the complexity of both the ensemble technique (boosting) and the base classifier (ANNs). RF is known to be faster in training than other machine learning algorithms (Benediktsson, Chanussot, and Fauvel 2007). ANNs are known to have a time-consuming training stage and boosting is a time consuming and computationally demanding algorithm (Benediktsson, Chanussot, and Fauvel 2007; DeFries and Chan 2000). The long training time is considered one of the disadvantages of the use of neural networks in classification of remotely sensed data (Pal and Mather 2003; Paola and Schowengerdt 1995), although studies have shown that ANNs outperform DTs in classification of remotely sensed data (Rogan et al. 2008). Our study showed that the use of ensembles of boosted ANNs does not outperform RF, while requiring more time for the classification process using the same training data. RF does not require the same computational effort needed by the boosting techniques (Benediktsson, Chanussot, and Fauvel 2007; Gislason, Benediktsson, and Sveinsson 2004, 2006; Horning 2010) and produced better results. However, the results showed that boosted ANNs have lower tendency for overfitting compared to RF. Boosting was also found to be robust to overfitting using other base classifiers (Polikar 2006; Ridgeway 1999; Schwenk and Bengio 2000). The impact of boosting algorithms was found to be more effective in cases of higher dimensional space of the input predictors (Canty 2010).

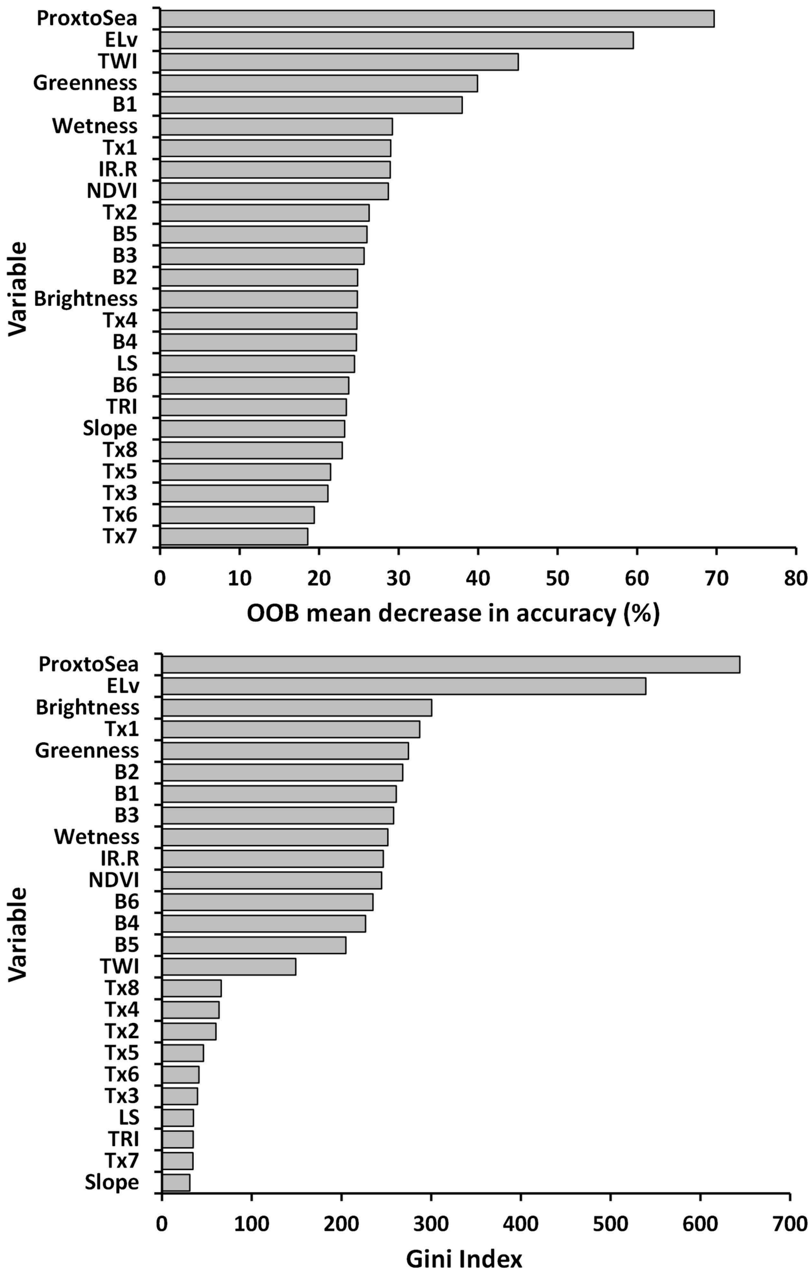

Figure 8. Variable importance contributions of different variables as measured by oob mean decrease in accuracy and the Gini Index (see Table 2 for variable abbreviation).

4. Conclusions

Using spectral and integrated multisource data, the two ensemble techniques performed well with respect to mapping heterogeneous desert landscapes located in the north-western coastal desert of Egypt. The two ensemble methods attained high accuracy metrics over three different data domains: spectral data; multisource data integrating spectral data with ancillary

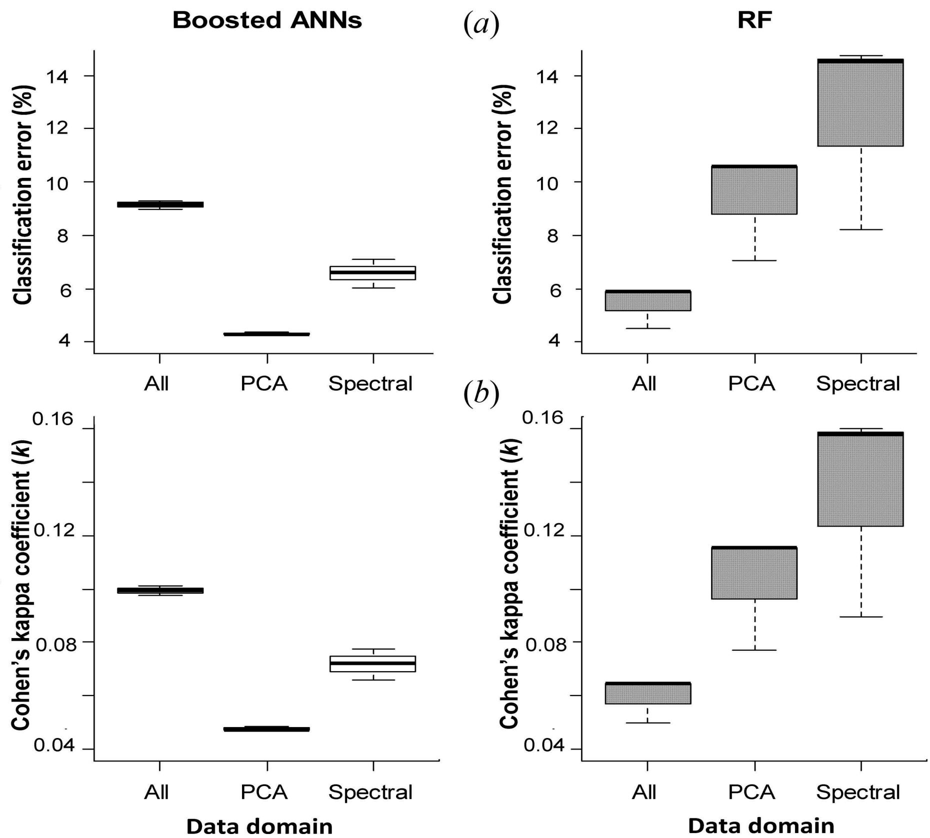

Figure 9. Difference between training accuracy and test accuracy of boosted ANNs and RF models using each of the data domains as measured by: (*a*) overall classification error and (*b*) Cohen's kappa coefficient (κ).

data; and transformed integrated data by PCA. In general, the overall accuracy exceeded 85% and overall κ attained values were over 0.83. The results indicate the merit of applying RF and boosted ANNs for classifying LULC in similar desert landscapes either to spectral data alone or using integrated spectral and ancillary data.

Inclusion of the ancillary data improved the classification performance of the two ensemble methods; however, the improvement in performance of RF exceeded that of the boosted ANNs. Integration of the ancillary data with spectral data resulted in improving the producer's accuracy of all the LULC classes in the study area using the two classifiers. This indicates that applying ensemble methods using integrated spectral and ancillary data achieves higher LULC classification accuracies for similar landscapes. Furthermore, this highlights the benefit of including ancillary data to help differentiate heterogeneous classes.

Random forests showed higher accuracy compared to boosted ANNs, but boosted ANNs showed better model generalization ability and lower overfitting tendencies when applied to the three different data domains. This study found random forests faster to train compared to the boosted ANNs that included the complexity of both the ensemble technique (boosting) and the base classifier (ANNs). This indicates that a compromise must be made when using these methods, taking into account the advantages and the limitations of each method. It should be noted that the performance of the two methods was different over the different data domains. This highlights the need for more testing of the two methods over different data domains bearing in mind that the results from each of these machine learning data-driven techniques is significantly influenced by the nature of the data being classified.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Alexandria University, Egypt [1/2/101- 2006/2007].

References

Abdel Razik, M. S. 2008. "Plant Diversity Changes in Response to Environmental Drivers and Pressures at El Omayed 'ROSELT/OSS' Observatory, Egypt." In *The Future of the Drylands*, Conference on Desertification and Drylands Research, Tunisia, edited by C. Lee and T. Schaaf, 19–21 June 2006, 289–309. France: UNESCO. Springer, Netherlands.

Aitkenhead, M. J., and I. H. Aalders. 2008. "Classification of Landsat Thematic Mapper Imagery for Land Cover Using Neural Networks." *International Journal of Remote Sensing* 29: 2075–2084. doi:10.1080/01431160701373739.

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer. 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*. Washington, DC: US Geological Survey Professional Paper 964.

Ayyad, M. A., and S. I. Ghabbour. 1993. "Dry Coastal Ecosystem of Eastern North Africa." In *Dry Coastal Ecosystems: Africa, America, Asia and Oceania*. Ecosystems of the World, edited by E. Van Der Maarel, 149–201. Vol. 2B. Amsterdam: Elsevier.

Ayyad, M. A., and E. Le Floc'H. 1983. *An Ecological Assessment of Renewable Resources for Rural Agriculture Development in the Western Mediterranean Coastal Region of Egypt*. Montpellier: UN Alexandria and C.N.R.S/C.E.P.E.L. Emberger.

Ball, J. 1939. *Contribution to the Geography of Egypt*, 308. Cairo: Survey and Mines Department.

Benediktsson, J. A., J. Chanussot, and M. Fauvel. 2007. "Multiple Classifier Systems in Remote Sensing: From Basics to Recent Developments." In *Mcs 2007*, LNCS, edited by M. Haindl, J. Kittler, and F. Roli, 501–512. Vol. 4472. Berlin: Springer-Verlag.

Bischof, H., W. Schneider, and A. J. Pinz. 1992. "Multispectral Classification of Landsat-Images Using Neural Networks." *IEEE Transactions on Geoscience and Remote Sensing* 30: 482–490. doi:10.1109/36.142926.

Breiman, L. 1996. "Bagging Predictors." *Machine Learning* 24: 123–140. doi:10.1007/BF00058655.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32. doi:10.1023/A:1010933404324.

Briem, G. J., J. A. Benediktsson, and J. R. Sveinsson. 2002. "Multiple Classifiers Applied to Multisource Remote Sensing Data." *IEEE Transactions on Geoscience and Remote Sensing* 40: 2291–2299. doi:10.1109/TGRS.2002.802476.

Camps-Valls, G. 2009. "Machine Learning in Remote Sensing Data Processing." In *IEEE International Workshop on Machine Learning for Signal Processing, 2009*. Grenoble: MLSP 2009.

Canty, M. J. 2009. "Boosting a Fast Neural Network for Supervised Land Cover Classification." *Computers & Geosciences* 35: 1280–1295. doi:10.1016/j.cageo.2008.07.004.

Canty, M. J. 2010. *Image Analysis, Classification, and Change Detection in Remote sensing: With Algorithms for ENVI/IDL*, 441. 2nd ed. Boca Raton, FL: Taylor & Francis Group, LLC.

Castaings, T., B. Waske, J. Atli Benediktsson, and J. Chanussot. 2010. "On the Influence of Feature Reduction for the Classification of Hyperspectral Images Based on the Extended

Morphological Profile." *International Journal of Remote Sensing* 31: 5921–5939. doi:10.1080/01431161.2010.512313.

Chan, J. C., and D. Paelinckx. 2008. "Evaluation of Random Forest and Adaboost Tree-Based Ensemble Classification and Spectral Band Selection for Ecotope Mapping Using Airborne Hyperspectral Imagery." *Remote Sensing of Environment* 112: 2999–3011. doi:10.1016/j.rse.2008.02.011.

Chander, G., B. L. Markham, and D. L. Helder. 2009. "Summary of Current Radiometric Calibration Coefficients for Landsat MSS, TM, ETM_+, and EO-1 ALI Sensors." *Remote Sensing of Environment* 113: 893–903. doi:10.1016/j.rse.2009.01.007.

Congalton, R. 1991. "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data." *Remote Sensing of Environment* 37: 35–46. doi:10.1016/0034-4257(91)90048-B.

Congalton, R., and K. Green. 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. 2nd ed. Boca Raton, FL: Taylor & Francis Group, LLC.

Cunningham, P., J. Carney, and S. Jacob. 2000. "Stability Problems with Artificial Neural Networks and the Ensemble Solution." *Artificial Intelligence in Medicine* 20: 217–225. doi:10.1016/S0933-3657(00)00065-8.

De Cosson, A. 1935. *Mareotis: A Short Account of the History and Ancient Monuments of the North Western Desert of Egypt and of Lake Mareotis*. London: Country Life.

DeFries, R. S., and J. C. Chan. 2000. "Multiple Criteria for Evaluating Machine Learning Algorithms for Land Cover Classification from Satellite Data." *Remote Sensing of Environment* 74: 503–515. doi:10.1016/S0034-4257(00)00142-5.

Dietterich, T. G. 1998. "Approximate Statistical Test for Comparing Supervised Classification Learning Algorithms." *Neural Computation* 10: 1895–1923. doi:10.1162/089976698300017197.

Dietterich, T. G. 2000. "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization." *Machine Learning* 40: 139–157. doi:10.1023/A:1007607513941.

Exelis Visual Information Solutions. 2010. Boulder, CO.

Foody, G. M. 2004. "Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy." *Photogrammetric Engineering & Remote Sensing* 70: 627–633. doi:10.14358/PERS.70.5.627.

Franklin, J. 2009. *Mapping Species Distribution: Spatial Inference and Prediction*, 320. New York: Cambridge University Press.

Franklin, S. E., R. J. Hall, L. M. Moskal, A. J. Maudie, and M. B. Lavigne. 2000. "Incorporating Texture into Classification of Forest Species Composition from Airborne Multispectral Images." *International Journal of Remote Sensing* 21: 61–79. doi:10.1080/014311600210993.

Freund, Y., and R. E. Schapire. 1996. "Experiments with a New Boosting Algorithm." In *Machine Learning, Proceedings of the Thirteenth International Conference (ICML '96)*, edited by L. Saitta, 148–56. Bari: Morgan Kaufmann Publishers.

Friedl, M. A., D. K. Mciver, J. C. F. Hodges, X. Y. Zhang, D. Muchoney, and A. H. Strahler. 2002. "Global Land Cover Mapping from MODIS: Algorithms and Early Results." *Remote Sensing of Environment* 83: 287–302. doi:10.1016/S0034-4257(02)00078-0.

Gessler, P. E., O. A. Chadwick, F. Chamran, L. Althouse, and K. Holmes. 2000. "Modeling Soil–Landscape and Ecosystem Properties Using Terrain Attributes." *Soil Science Society of America Journal* 64: 2046–2056. doi:10.2136/sssaj2000.6462046x.

Gessler, P. E., I. D. Moore, N. J. McKenzie, and P. J. Ryan. 1995. "Soil-Landscape Modelling and Spatial Prediction of Soil Attributes." *International Journal of Geographical Information Systems* 9: 421–432. doi:10.1080/02693799508902047.

Gislason, P., J. Benediktsson, and J. Sveinsson. 2006. "Random Forests for Land Cover Classification." *Pattern Recognition Letters* 27: 294–300. doi:10.1016/j.patrec.2005.08.011.

Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson. 2004. "Random Forest Classification of Multisource Remote Sensing and Geographic Data. Geoscience and Remote Sensing Symposium, 2004. IGARSS 04." *Proceedings. IEEE International* 2: 1049–1052.

Gong, P., D. J. Marceau, and P. J. Howarth. 1992. "A Comparison of Spatial Feature Extraction Algorithms for Land-Use Classification with SPOT HRV Data." *Remote Sensing of Environment* 40: 137–151.

Halmy, M. W., P. Gessler, J. A. Hicke, and B. B. Salem. 2015. "Land Use/Land Cover Change Detection and Prediction in the North-Western Coastal Desert of Egypt Using Markov-Ca." *Applied Geography* 63: 101–112. doi:10.1016/j.apgeog.2015.06.015.

Halmy, M. W. A. 2012. “Environmental Risk Assessment for Habitats of the Egyptian Northwestern Coastal Desert.” Doctoral dissertation, University of Idaho, USA, ProQuest Dissertations and Theses.
Haralick, R. M., K. Shanmugam, and I. Dinstein. 1973. “Textural Features for Image Classification.” *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3: 610–621. doi:10.1109/TSMC.1973.4309314.
He, L., F. Kong, and Z. Shen. 2006. “Artificial Neural Network Ensemble for Land Cover.” In *Proceedings of the 6th World Congress on Intelligent Control and Automation*, June 21–23 2006, 10054–10057. Dalian: IEEE.
Horning, N. 2010. “Random Forests : An Algorithm for Image Classification and Generation of Continuous Fields Data Sets.” In *International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Applied Sciences Sciences*, New York.
Ismail, R. R., and O. O. Mutanga. 2011. “Discriminating the Early Stages of Sirex Noctilio Infestation Using Classification Tree Ensembles and Shortwave Infrared Bands.” *International Journal Of Remote Sensing* 32: 4249–4266. doi:10.1080/01431161.2010.486413.
Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara, 2008. “Hole-Filled SRTM for the Globe Version 4, Available from the CGIAR-CSI SRTM 90m Database.” Accessed September 21, 2011. http://srtm.csi.cgiar.org
Kauth, R. J., and G. S. Thomas. 1976. “The tasseled Cap: A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by LANDSAT.” In *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*, 4B-41–4B-51. Indiana: Purdue University of West Lafayette.
Kavzoglu, T., and P. M. Mather. 1999. “Pruning Artificial Neural Networks: An Example Using Land Cover Classification of Multi-Sensor Images.” *International Journal of Remote Sensing* 20: 2787–2803. doi:10.1080/014311699211796.
Kavzoglu, T., and P. M. Mather. 2003. “The Use of Back-Propagating Artificial Neural Networks in Land Cover Classification.” *International Journal of Remote Sensing* 24: 4907–4938. doi:10.1080/0143116031000114851.
Kiema, J. B. K. 2002. “Texture Analysis and Data Fusion in the Extraction of Topographic Objects from Satellite Imagery.” *International Journal of Remote Sensing* 23: 767–776. doi:10.1080/01431160010026005.
Kotsiantis, S. B. 2007. “Supervised Machine Learning: A Review of Classification Techniques.” *Informatica* 31: 249–268.
Liaw, A., and M. Wiener. 2002. “Classification and Regression by Randomforest.” *R News* 2: 18–22.
Lu, D., S. Hetrick, and E. Moran. 2010. “Land Cover Classification in a Complex Urban-Rural Landscape with QuickBird Imagery.” *Photogrammetric Engineering & Remote Sensing* 76: 1159–1168. doi:10.14358/PERS.76.10.1159.
Maier, H. R., and G. C. Dandy. 2000. “Neural Networks for the Prediction and Forecasting of Water Resources Variables : A Review of Modelling Issues and Applications.” *Environmental Modelling & Software* 15: 101–124. doi:10.1016/S1364-8152(99)00007-9.
Mas, J. F. 2004. “Mapping Land Use/Cover in a Tropical Coastal Area Using Satellite Sensor Data, GIS and Artificial Neural Networks.” *Estuarine, Coastal and Shelf Science* 59: 219–230. doi:10.1016/j.ecss.2003.08.011.
Mas, J. F., and J. J. Flores. 2008. “The Application of Artificial Neural Networks to the Analysis of Remotely Sensed Data.” *International Journal of Remote Sensing* 29: 617–663. doi:10.1080/01431160701352154.
McInerney, D. O., and M. Nieuwenhuis. 2009. “A Comparative Analysis of K NN and Decision Tree Methods for the Irish National Forest Inventory.” *International Journal of Remote Sensing* 30: 4937–4955. doi:10.1080/01431160903022936.
Miao, X., J. S. Heaton, S. Zheng, D. Charlet, and H. Liu. 2012. “Applying Tree-Based Ensemble Algorithms to the Classification of Ecological Zones Using Multi-Temporal Multi-Source Remote-Sensing Data.” *International Journal of Remote Sensing* 33: 1823–1849. doi:10.1080/01431161.2011.602651.
Moody, J. 1991. “Note on Generalization, Regularization, and Architecture Selection in Nonlinear Learning Systems.” In *Proceedings of the First IEEE Workshop on Neural Networks for Signal Processing*, edited by B. H. Juang, S. Y. Kung, and C. A. Kamm, 1–10. Piscataway, New Jersey: IEEE.

Moore, I. D., P. E. Gessler, G. A. Nielsen, and G. A. Peterson. 1993. "Soil Attribute Prediction Using Terrain Analysis." *Soil Science Society of America Journal* 57: 443–452. doi:10.2136/sssaj1993.03615995005700020026x.

Olaya, V. 2009. "Basic Land-Surface Parameters." In *Geomorphometry*, edited by T. Hengl and H. I. Reuter, 141–169. Vol. 33. Oxford: Elsevier Science.

Pal, M. 2005. "Random Forest Classifier for Remote Sensing Classification." *International Journal of Remote Sensing* 26: 217–222. doi:10.1080/01431160412331269698.

Pal, M. 2006. "Support Vector Machine Based Feature Selection for Land Cover Classification: A Case Study with DAIS Hyperspectral Data." *International Journal of Remote Sensing* 27: 2877–2894. doi:10.1080/01431160500242515.

Pal, M., and P. M. Mather. 2003. "An Assessment of the Effectiveness of Decision Tree Methods for Land Cover Classification." *Remote Sensing of Environment* 86: 554–565. doi:10.1016/S0034-4257(03)00132-9.

Paola, J. D., and R. A. Schowengerdt. 1995. "A Review and Analysis of Backpropagation Neural Networks for Classification of Remotely Sensed Multi-Spectral Imagery." *International Journal of Remote Sensing* 16: 3033–3058. doi:10.1080/01431169508954607.

Pijanowski, B., D. Brown, B. Shellito, and G. Manik. 2002. "Using Neural Networks and GIS to Forecast Land Use Changes: A Land Transformation Model." *Computers, Environment and Urban Systems* 26: 553–575. doi:10.1016/S0198-9715(01)00015-1.

Polikar, R. 2006. "Ensemble Based Systems in Decision Making." *IEEE Circuits And Systems Magazine* 6: 21–45. doi:10.1109/MCAS.2006.1688199.

Pontius, R. G., and M. Millones. 2011. "Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment." *International Journal of Remote Sensing* 32: 4407–4429. doi:10.1080/01431161.2011.552923.

R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*. Vienna: The R Foundation for Statistical Computing. ISBN:3-900051-07-0.

Ridd, M. K. 1995. "Exploring a V-I-S (Vegetation-Impervious Surface-Soil) Model for Urban Ecosystem Analysis through Remote Sensing: Comparative Anatomy for Cities." *International Journal of Remote Sensing* 16: 2165–2185. doi:10.1080/01431169508954549.

Ridgeway, G. 1999. "The State of Boosting." *Computing Science and Statistics* 31: 172–181.

Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. 2012. "An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover Classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 67: 93–104. *International Society for Photogrammetry and Remote Sensing*, Inc. (ISPRS). doi:10.1016/j.isprsjprs.2011.11.002.

Rogan, J., J. Franklin, D. Stow, J. Miller, C. Woodcock, and D. Roberts. 2008. "Mapping Land-Cover Modifications over Large Areas: A Comparison of Machine Learning Algorithms." *Remote Sensing of Environment* 112: 2272–2283. doi:10.1016/j.rse.2007.10.004.

Rokach, L. 2005. "Ensemble Methods for Classifiers." In *The Data Mining and Knowledge Discovery Handbook*, edited by O. Maimon and L. Rokach, 957–980. New York: Springer-Verlag.

ROSELT/OSS. 2005. "El Omayed ROSELT/OSS Observatory (Scientific report Part 1)." Faculty of Science - University of Alexandria, Egypt. Accessed July 21, 2008. http://prog.oss.org.tn/roselt/index.php?option=com_jotloader&view=categories&cid=5_9cd2e6a5def039b9a1ef361919965964&Itemid=78&lang=en

Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. "Learning Internal Representations by Error Propagation." In *In Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Foundations, edited by D. E. Rumelhart and J. L. McClelland, 318–362. Cambridge, MA: MIT Press.

SAGA. 2011. "System for Automated Geoscientific Analyses." Accessed February 13. http://www.saga-gis.org

Said, R. 1962. *The Geology of Egypt*. Amsterdam: Elsevier.

Salem, B. B. 1989. "Remote Sensing of Vegetation and Land Use in the North Western Desert of Egypt." Doctoral dissertation. Alexandria University, Alexandria, Egypt, 244.

Salem, B. B. 2003a. "Assessing Habitat Fragmentation and Its Effect on Plant Biodiversity Using Multidate Satellite Imagery. Case Study: Omayed Biosphere Reserve (OBR), Western Coastal Desert of Egypt." *Egyptian Journal of Desert Research* 53: 1–17.

Salem, B. B. 2003b. "Biosphere Reserves on North-Western Egyptian Coast, a Site for Monitoring Biodiversity and Integrated Water Management." *Developments in Water Science* 50: 119–128.
Salem, B. B. 2005. "Forms of Commonness and Rarity of the Threatened and Endangered Perennial Species in the Western Coastal Desert of Egypt: Implications for Conservation and Management." *Assuit University Journal of Botany* 34: 289–310.
Schwenk, H., and Y. Bengio. 2000. "Boosting Neural Networks." *Neural Computation* 12: 1869–1887. doi:10.1162/089976600300015178.
Serpico, S. B., L. Bruzzone, and F. Roli. 1996. "An Experimental Comparison of Neural and Statistical Non-Parametric Algorithms for Supervised Classification of Remote-Sensing Images." *Pattern Recognition Letters* 17: 1331–1341. doi:10.1016/S0167-8655(96)00090-6.
Sesnie, S. E., B. Finegan, P. E. Gessler, S. Thessler, Z. R. Bendana, and A. M. S. Smith. 2010. "The Multispectral Separability of Costa Rican Rainforest Types with Support Vector Machines and Random Forest Decision Trees." *International Journal of Remote Sensing* 31: 2885–2909. doi:10.1080/01431160903140803.
Sesnie, S. E., P. E. Gessler, B. Finegan, and S. Thessler. 2008. "Integrating Landsat TM and SRTM-DEM Derived Variables with Decision Trees for Habitat Classification and Change Detection in Complex Neotropical Environments." *Remote Sensing of Environment* 112: 2145–2159. doi:10.1016/j.rse.2007.08.025.
Sluiter, R., and E. J. Pebesma. 2010. "Comparing Techniques for Vegetation Classification Using Multi- and Hyperspectral Images and Ancillary Environmental Data." *International Journal of Remote Sensing* 31: 6143–6161. doi:10.1080/01431160903401379.
Stuckens, J., P. R. Coppin, and M. E. Bauer. 2000. "Integrating Contextual Information with Per-Pixel Classification for Improved Land Cover Classification." *Remote Sensing of Environment* 71: 282–296. doi:10.1016/S0034-4257(99)00083-8.
Tong, X., X. Zhang, and M. Liu. 2010. "Detection of Urban Sprawl Using A Genetic Algorithm-Evolved Artificial Neural Network Classification in Remote Sensing: A Case Study in Jiading and Putuo Districts of Shanghai, China." *International Journal of Remote Sensing* 31: 1485–1504. doi:10.1080/01431160903475290.
Waske, B., J. A. Benediktsson, K. Árnason, and J. R. Sveinsson. 2009. "Mapping of Hyperspectral AVIRIS Data Using Machine-Learning Algorithms." *Canadian Journal of Remote Sensing* 35: S106–116. doi:10.5589/m09-018.
Weng, Q. 2002. "Land Use Change Analysis in the Zhujiang Delta of China Using Satellite Remote Sensing, GIS and Stochastic Modelling." *Journal of Environmental Management* 64: 273–284. doi:10.1006/jema.2001.0509.
Yool, S. R. 1998. "Land Cover Classification in Rugged Areas Using Simulated Moderate-Resolution Remote Sensor Data and an Artificial Neural Network." *International Journal of Remote Sensing* 19: 85–96. doi:10.1080/014311698216440.
Zhang, G. P. 2000. "Neural Networks for Classification: A Survey." *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews* 30: 451–462. doi:10.1109/5326.897072.