Journal of
Applied Remote Sensing

# Supervised machine learning of fused RADAR and optical data for land cover classification

Guido Cervone
Barry Haack

# Supervised machine learning of fused RADAR and optical data for land cover classification

**Guido Cervone and Barry Haack**
George Mason University, Department of Geography and Geoinformation Science, 4400 University Drive, MS 6C3, Fairfax, Virginia 22030-4444
{gcervone,bhaack}@gmu.edu

**Abstract.** Supervised machine learning algorithms are used to classify pixels of a multi-sensor remote sensing dataset comprising RADAR and optical measurements for central Sudan. A total of 19 layers were used, 16 RADAR bands from RADARSAT DN, and texture bands acquired on 13 December 2008 (dry season) and on 2 June 2009 (wet season), and three optical bands acquired by Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) on 25 February 2009. Three different machine learning supervised classification algorithms were used to test the advantage of combining RADAR and optical data: a decision rule, a decision tree, and a naive Bayesian. In all the experiments performed, a combination of RADAR and optical bands leads to higher predictive accuracy and better land cover classification than either sensor used independently. The decision rule classifier performed best among the three methods used. © *2012 Society of Photo-Optical Instrumentation Engineers (SPIE).* [DOI: 10.1117/1.JRS.6.063597]

**Keywords:** artificial intelligence; pixel classification; remote sensing; rule classifiers.

Paper 12089 received Apr. 2, 2012; revised manuscript received Sep. 24, 2012; accepted for publication Oct. 4, 2012; published online Oct. 30, 2012.

## 1 Introduction

Formulating general hypotheses from limited observations is one of the fundamental principles of scientific discovery. As more spaceborne remote sensing data become available due to a higher number of sensors with increased spatial and spectral resolutions, better storage and sharing infrastructure, and deeper understanding of the physical processes, analyzing data to generate hypotheses becomes a crucial and difficult task in most problem solving applications. An improved ability to extract accurate information from remote sensing can be accomplished by two often separate approaches: 1. improved processing strategies such as those provided by machine learning and 2. improved data sets such as multi-temporal or multi-sensor. This study integrates these two approaches by employing different machine learning classifiers with data from RADAR and optical sensors.

The proposed methodology is based on machine learning classifiers that develop general hypotheses from labeled pixels, and uses the learned knowledge to classify unknown pixels. Three different classifiers are tested: a decision rule induction algorithm, a decision tree learner, and a naive Bayesian classifier. The data used in this study are a fusion of RADAR and optical measurements.

The decision tree and naive Bayesian classifiers were used because they have been extensively used in the literature (e.g., Refs. 1 and 2). The decision rule classifier was chosen because it is often under-represented when classifiers are compared, and it has the theoretical advantage of a richer representation language, which can lead to more sophisticated partitioning of the solutions pace, and also a higher understandability of the learned knowledge as rules are usually easier to inspect and validate than trees and probabilities. Therefore, the contributions of the article can be summarized as follows:

0091-3286/2012/$25.00 © 2012 SPIE

(1) A theoretical analysis of the strengths and weaknesses of three different classifiers, one of which has often been under-represented.
(2) The generation and analysis of a fused RADAR and optical dataset.
(3) A performance analysis of the classifiers over a problem of high complexity.
(4) A discussion of how the fused data affects the predictive accuracy of pixel classification.

### 1.1 *Land Cover Classification*

Supervised learning consists of analyzing examples and counter-examples for each land cover class, and creating generalizations of the data to learn concepts. These concepts describe patterns in attribute space, and can be in the form of statistical summarizations, probabilities, network graphs, or inductive rules, to classify unknown pixels. For the case of remote sensing land cover classification, generally each pixel is described by a vector of attributes corresponding to the reflectivity of the pixel in different parts of the electromagnetic (EM) spectrum. The attribute space can be viewed as a cube, where the vertical and horizontal dimensions correspond to the spatial resolution, and the depth is the spectral resolution of the data.

Statistics and machine learning techniques have been used extensively for the supervised land cover classification problem of remote sensing data (e.g., Refs. 3 to 7). Rogan et al.[1] have tested different machine learning methods for mapping land cover modifications. They have shown that the overall methodology is sound, and different algorithms present advantages and disadvantages depending on the type, quality, and size of data used. One of the algorithms they investigated generated decision rules, but only as a post-processing operation for the decision tree learned. In fact, although it is possible to transform decision trees to decision rules, they usually use a weaker representation language than when decision rule algorithms are employed. For example, rules converted from trees normally cannot include internal disjunctions, because these are not representable in a tree structure.

A different line of research concentrates on augmenting the spectral information of each pixel with additional data. Walter[8] described an object-based change detection approach for the classification of remote sensing multispectral data. Instead of classifying single pixels, it acts on groups of pixels that represent already existing objects in a geographical information system (GIS) database. Texture information was also investigated as a mean to improve the accuracy of land cover classification algorithms.[9–11] Puig and García[12] explored different ways for extracting texture information for image classification. They concluded that pixel-based texture classification can be significantly improved by evaluating a given texture method over multiple windows of different size and then by integrating the results through a classical Bayesian scheme.

### 1.2 *RADAR Data Fusion*

Central to the land cover classification problem are the quality and characteristics of the remote sensing data used. Traditional means of providing reliable land cover/use information via remote sensing has been primarily undertaken by multispectral systems such as Landsat, SPOT, Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), and IRS AWiFS. These systems are not able to fully satisfy the demands for providing land cover/use information for some areas around the world. However, with the availability of operational spaceborne RADAR systems, these limitations may be removed.

RADAR systems have some advantages over traditional multispectral sensors. The longer wavelengths of RADAR are capable of penetrating atmospheric conditions that limit traditional spaceborne optical and multispectral systems.[13] This longer wavelength operational characteristic holds enormous data collecting potential for many geographic areas around the world that are often obscured by cloud cover or areas of high latitudes with limited daylight.

The surface interaction of RADAR is very different than optical data, thus providing unique information about ground features. The response of RADAR is more a function of surface roughness, geometry, and internal structure, as opposed to surface reflection with optical wavelengths. The variation in RADAR backscatter from a feature may be a result of incident angle, look

direction, moisture on the surface, date, or the physical composition of the feature itself. Backscatter is also strongly influenced by orientation of the feature to the incoming RADAR signal.

Conversely, one of the difficulties with the analysis of RADAR data is that most recent RADAR spaceborne systems only collected data using a single wavelength with a fixed polarization. Hence only one component of the total surface scattering is thereby being measured, while any additional information contained within the returned RADAR signal is lost.[14,15] More recent systems, such as the Japanese ALOS PALSAR and the Canadian RADARSAT-2, include an increased number of polarizations. Imagery acquired under different polarizations will obtain different backscatter responses and different informational content.[16–19]

Textural information may be as important as spectral information in RADAR, as the information content of an image resides in both the intensity (spectral) of individual pixels and the spatial arrangement of the pixels.[20,21] Standard image classification procedures used to extract information from remotely sensed images usually ignore this spatial information and are based on purely spectral characteristics. Such classifiers will be ineffective when applied to land cover/use classes such as residential and urban areas that are largely distinguished by their spatial rather than their spectral characteristics.[22,23] The advantages of using derived RADAR measures, such as texture measures at different window sizes, in comparison to original RADAR data have been demonstrated by Haack et al.[24] and Herold et al.[25] Textural information may be used in combination with the spectral measurements of a wavelength for analysis.[26,27] Texture is particularly useful because it indicates the local variability of gray level in the spatial domain, revealing unique information about the ground feature.[28]

The availability of remotely sensed data for the same geographic area obtained from separate sensors, operating in different portions of the electromagnetic spectrum, such as Landsat and RADARSAT, has increased greatly. This, along with improved technology for the processing and fusing of such separate data sets, has made the synergies between optical and RADAR data for land applications of greater practical importance.[29–31] The fusing of data from different sensors is done in an attempt to generate an interpretation of a geographic area that is not obtainable from any single sensor alone and also to reduce the uncertainty associated with data from a single source.[32–34]

## 2 Study Site and Data

The study site as described in the following was in central Sudan, a rapidly expanding area, familiar to the authors, who contains distinct land cover types, and thus is particularly suited for testing the proposed methodology.

The data from both sensors were acquired at different spatial resolutions and fused by georectifying and resampling the pixels to a common 15 m grid. The integrated data consist of 19 bands with each band containing 2103 columns by 2615 rows of pixels. The full data set consisted of three ASTER and two dates of RADARSAT with derived texture. Different datasets were created to individually test ASTER data (A09), 2008 RADARSAT data (R08), and 2009 RADARSAT data (R09), and their combinations [(R08 + A09 and R09 + A09); for example, see Table 1].

### 2.1 *Wad Madani*

The study site includes the city of Wad Madani, the capital of the Al Jazirah state in east-central Sudan. Wad Madani is located on the west bank of the Blue Nile River, approximately 136 km southeast of the Sudanese capital city, Khartoum. The city is the center of the Al Jazirah cotton-growing region which results from one of the largest agriculture irrigation projects in the world, started by the British in 1925. The waters of the Blue Nile River are currently directed through more than 4300 km of irrigation canals and ditches. With Wad Madani as its hub, this region is an important world cotton producer, in addition to prosperous local trade in wheat, peanuts, barley, and livestock. Regional populations surround the banks of the White and Blue Nile rivers, with a population concentration of more than 200,000 people in Wad Madani.

**Table 1** Description of the 19 bands used in the study. H and V indicate RADAR horizontal and vertical polarization.

| Layer ID | Dataset ID | Name | Description |
|---|---|---|---|
| 1 | R08 | RADARSATDec08.b1 | HH |
| 2 | R08 | RADARSATDec08.b2 | HV |
| 3 | R08 | RADARSATDec08.b3 | VH |
| 4 | R08 | RADARSATDec08.b4 | VV |
| 5 | R08 | RADARSATDec08.text.b5 | Texture HH |
| 6 | R08 | RADARSATDec08.text.b6 | Texture HV |
| 7 | R08 | RADARSATDec08.text.b7 | Texture VH |
| 8 | R08 | RADARSATDec08.text.b8 | Texture VV |
| 9 | A09 | Aster.b9 | Visible (Green) |
| 10 | A09 | Aster.b10 | Visible (Red) |
| 11 | A09 | Aster.b11 | Infrared |
| 12 | R09 | RadSatJun09.b12 | HH |
| 13 | R09 | RadSatJun09.b13 | HV |
| 14 | R09 | RadSatJun09.b14 | VH |
| 15 | R09 | RadSatJun09.b15 | VV |
| 16 | R09 | RadSatJun09.text.b16 | Texture HH |
| 17 | R09 | RadSatJun09.text.b17 | Texture HV |
| 18 | R09 | RadSatJun09.text.b18 | Texture VH |
| 19 | R09 | RadSatJun09.text.b19 | Texture VV |

The diverse range of land covers present in the study area have relatively distinct delineations between them. Most of the lands to the west of the Blue Nile are agricultural areas, presumably for cotton or another one of the local crops. The city of Wad Madani has urban features occupying a peninsula-like area formed by a bend in the Blue Nile. North of Wad Madani and east of the Blue Nile, land cover tends to often be bare soil with sparse natural shrubs/trees.

## 2.2 *RADARSAT*

The first Canadian RADARSAT system has been operational since 1995. It was a C-band, 5.7 cm, single wavelength, and single polarization, horizontal-horizontal (HH), system. RADARSAT-2 was launched in December 2007 and was a quad polarization system [HH, horizontal-vertical (HV), vertical-horizontal (VH), and vertical-vertical (VV)]. Both sensors can collect data at various depression angles, spatial resolutions, and swath widths. RADARSAT can collect data with spatial resolutions ranging from about $1 \times 2$ m to $100 \times 160$ m and with a scene size from 8 by 18 km to 500 by 500 km. The sensor can also collect data in various incidence angles.

The data for this study are fine quad-pol single look complex (SLC) with a spatial resolution of 5.4 by 8.0 m and a scene size of 25 by 25 km. The RADARSAT data were originally collected as 11 bit data but were compressed to 8 bit for this analysis. The first image data, 13 December 2008, were collected during the dry season when most crops are dormant. The second data, 2 June 2009, were collected during the wet season with the crops in full development. One aspect

of this study was to examine the relative value of the RADAR data at the different seasons. One difficulty, however, with this analysis is that the ground conditions will be seasonally different for some classes creating some issues with the comparisons. From the two dates of original RADAR bands, eight additional bands were created based upon their variance texture over an $11 \times 11$ pixel window.

Each of the RADARSAT quad polarization bands will generally have different amounts of backscatter from the same surface feature as a function of the feature-scattering mechanisms. Some general observations can be made about the various polarizations.

The like polarization HH band is more capable of penetrating vegetation, particularly crops, for access to the underlying soil and has been employed for soil moisture studies. The HH also is more suitable for separating water from ice. The VV band is sensitive to water roughness and thus more useful for studies of wave height and action. Cross-polarization bands, HV and VH, are generally weaker in backscatter than like polarization.

These cross-polarization bands are often useful for vegetation discrimination and in particular for separation of broad-leaved crops from grain crops since the geometries of broad-leaf vegetation cause multiple-bounce scattering, resulting in some cross-polarized backscatter. Since the HV backscatter from water surfaces is much reduced compared to the VV backscatter, the HV channel is very suitable for detecting features on the water surface that create multiple scattering. Examples of such targets are ship superstructures and ice deformations (ridging, fractures, and rubble).

## 2.3 *ASTER*

ASTER is an imaging instrument flying on NASA Terra. ASTER is a cooperative research effort between NASA, Japan's Ministry of Economy, Trade and Industry (METI), and Japan's Earth Remote Sensing Data Analysis Center (ERSDAC). The principal mission of ASTER is to collect high spatial resolution data of land surface temperature, reflectance, and elevation. ASTER collected 12 spectral channels in three general regions at varied spatial resolutions. Three bands of visible-near infrared data were collected at 15 m pixel size, six bands in the short wave infrared at 30 m, and five thermal bands at 90 m. The system had a 60 km swath and pointable sensors capable of providing stereo coverage. For this study, three bands for visible green, visible red, and near infrared at 15 m were obtained on 25 February 2009.

## 2.4 *Training and Testing Regions*

Twenty-one different regions were selected from the data and classified into four different land cover types: bare soil, agriculture, urban, and water. From these, 11 regions were used for training the classifier (calibration), and the remaining 10 regions were used for testing the learned hypotheses (validation). The regions were selected because it was possible to clearly distinguish between the different cover types, thus not introducing significant noisy pixels. A different number of regions was selected for training and testing in order to keep a 1:2 ratio between the number of training and testing pixels. For example, there are three regions for the training set of agriculture (1865 pixels total) and two regions for bare soil (1485 pixels total).

Table 2 summarizes the number of pixels for the training and testing regions used, and their cumulative sizes. Notice that in some cases (e.g., Training-Agriculture) there are three regions used, while in other cases (e.g., Testing-Urban) there are two regions used. Figure 1(a) shows the

**Table 2** Number of pixels for the training and testing regions used (either 2 or 3 for each class).

| Type | Agriculture | Bare soil | Urban | Water | Total |
|---|---|---|---|---|---|
| Training | $579 + 612 + 674 = 1865$ | $560 + 925 = 1485$ | $608 + 1121 + 1002 = 2731$ | $507 + 757 + 170 = 1434$ | 7515 |
| Testing | $809 + 2268 + 798 = 3875$ | $491 + 1567 + 2063 = 4121$ | $2040 + 1873 = 3913$ | $1628 + 1375 = 3003$ | 14,912 |

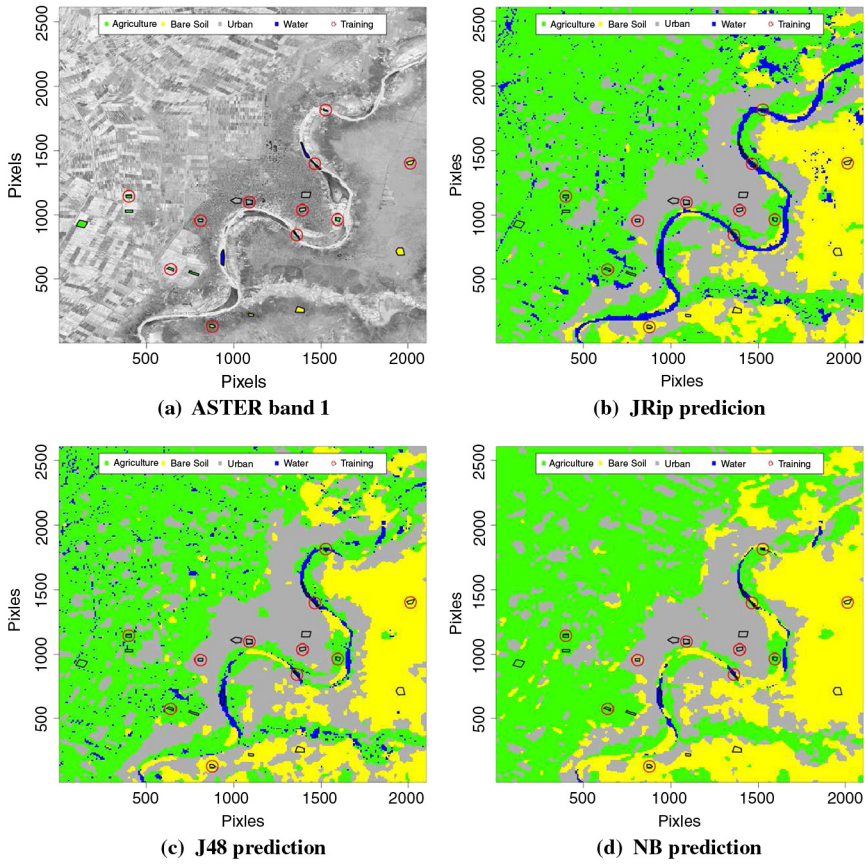



**Fig. 1** (a) Image of the Wad Madani area. The training (circled) and testing areas are shown for each of the four cover types, (b) scene prediction using JRip, (c) scene prediction using J48, and (d) scene prediction using Naive Bayesian.

study area, along with the training (indicated with an inscribing circle) and testing areas. Additionally, 62,500 pixels were uniformly selected from the original data for a final classification of the entire scene.

The training and testing data consist of the spectral values of the pixels of the selected regions. Each row represents a pixel, and is represented with 20 values, one for each of the 19 bands, plus the class information, indicating which region (e.g., urban) the pixel belongs to.

## 3 Methodology

Machine learning classifiers are used to learn hypotheses that discriminate the spectral characteristics of the four land cover classes. The input information includes labeled data or, in other words, data that are already assigned to a particular class or group. Unlike clustering, a form of unsupervised learning whose goal is dividing unlabeled data into distinct classes, in supervised learning, labeled data are generalized to identify the characteristics of the entire class.

In its simplest form, given two sets of multivariate descriptions, or events, $P_1 \ldots P_n$ and $N_1 \ldots N_m$, machine learning classifiers find rules that cover all $P$ examples (also known as positive events), and do not cover any of the $N$ examples (also known as negative events). More generally, each multivariate description is a classified event of type $x_1, \ldots x_k$, and $c$, where each $x$ is an attribute value, and $c$ is the class it belongs to. For each class $c$, positives are all the events that belong to class $c$, and negatives are all the events belonging to the other classes.

During the learning phase, all the pixels belonging to the class for which the hypotheses are being generated are the positives, and all pixels belonging to the remaining three classes are the negatives. The learned hypotheses are used to predict the class of the testing data. An estimation of accuracy is provided by comparing the predicted and truth value. The results are provided in

terms of predictive accuracy, a value indicating the percentage of correct predictions on the testing data, and using a confusion matrix, showing the overall summary of the class predictions for each of the land cover types. When a good model is identified, it is used to predict all pixels in the scene. The overall goal can be summarized as providing labels for all the pixels, while providing labels only for a small portion of the pixels through the training data.

Supervised machine learning classification is a general paradigm that can be implemented using different algorithms. In this paper, we have investigated three different general methodologies: decision rules, decision trees, and naive Bayesian. The learning algorithms were implemented using the Waikato Environment for Knowledge Analysis (WEKA) library (http://www.cs.waikato.ac.nz/ml/weka/). WEKA is a collection of machine learning algorithms for data mining tasks.[2,35–37] WEKA contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

### 3.1 *JRip: Decision Rules*

JRip is a propositional rule learner, and is the WEKA implementation of the repeated incremental pruning to produce error reduction (RIPPER), proposed by Cohen[38] as an optimized version of incremental reduced error pruning (IREP). The rules of RIPPER are used to classify an unknown event into one of the original training classes.

The learning process is based on an iterative loop that 1. generates new conditions, 2. prunes them, and 3. optimizes them. The process stops when all positive examples are covered, or when a large error rate is detected (usually $\geq$50%). During the learning process, different measures are used to evaluate the quality of the learned descriptions and are based on information gain and coverage of positive and negative examples. One of the main advantages of RIPPER is the highly descriptive rules that make it possible to inspect and validate the learned knowledge. RIPPER can learn rules from data that contain both numerical and categorical data. Discretization of continuous variables is automatically performed during the learning process, which makes RIPPER suitable for use with numerical data, such as the remote sensing data employed in this study.

Each RIPPER rule takes the form of

$$\text{conclusion} \leftarrow \text{consequent} : \text{annotation}, \tag{1}$$

where conclusion is the group class (e.g., urban), consequent is a conjunction of attribute relation conditions, and annotation is a summary of how many positive (p) and negative (n) events from the training set the rule covers. The annotation is not a measure of predictive ability of the rules, but how complete and consistent they are with regard to the training set. Each class can have (and usually does) multiple rules, called ruleset or cover.

The WEKA implementation of RIPPER differs slightly with respect to the original implementation. Most importantly, two documented bugs have been fixed that would affect the ruleset size and accuracy slightly.[35]

### 3.2 *J48: Decision Trees*

J48 is the WEKA implementation of the C4.5 trees induction algorithm, proposed by Quinlan[39] as an improved version of the earlier ID3 algorithm, developed by the same author. J48 learns decision trees from labeled data by using the concept of information entropy.[40]

The learning process consists in a progressive partition of the search space, where at each step a decision is made to divide the search space in smaller areas, and continues until each area contains only events of the same class. C4.5 is thus a "divide and conquer algorithm" because when each split is made, the search space is subdivided into smaller problems (divide) and each subproblem is solved individually (conquer). To decide which attribute-value combination to use for the split, C4.5 computes the entropy for each of the attributes and chooses the combination with the highest value. The leaves of the tree indicate the decision class for the entire area described by the repetitive splits. This first part of the algorithm is based on an inductive generalization step, where single instances are generalized to correspond with areas. This generalization

step is followed by a specialization step, called pruning, in which some areas are enlarged and some splits removed in order to simplify the tree, with the expenses of introducing some inconsistencies.

J48 can learn trees with data that contains numerical and categorical data, noise, and missing data. Numerical data do not require discretization.[41]

The classification of an unknown event consists in checking all conditions, starting from the root. The leaves of the tree indicate the predicted class for the unknown event. Although it is possible to convert the learned trees into rules, the resulting descriptions are usually expressed in simpler representation language. For example, C4.5 rules only allow for atomic relationships between attributes and possible values and do not allow for internal disjunctions or multiple ranges.

The main advantage of "divide-and-conquer" algorithms is that they are very fast, because at each step they divide the problem into smaller ones, thus always analyzing a smaller number of events. However, the main methodological disadvantage is that consecutive splits are made on a decreasing number of examples, which translates into decisions made with a smaller amount of information. In contrast, decision rule learners such as RIPPER consider all the search space at each iteration, making the algorithm considerably slower, but ensuring that all hypotheses are learned with the maximum amount of information available.

### 3.3 *NB: Naive Bayesian Classifier*

NB is the WEKA implementation of a classic naive Bayesian classifier. It is a probabilistic classifier which uses Bayes' theorem to assign events to classes.[40,42,43]

NB collects frequency counts of training events and builds statistics for each band. The classification of an unknown event is made by comparing the attribute values of the events with the statistics of each class. The class with the highest similarity is chosen.

The method is called "naive" because of the assumption that it is possible to build statistics for every attribute, and it assumes a full independence among all attributes. In practice, the methodology has proven to work well on a variety of problems, and it is considered a very effective supervised classifier.[43]

## 4 Results

Experiments were performed to test different combinations of RADAR and optical data for each of the three machine learning classifiers used. The learned hypotheses were used to predict the land cover class for the entire scene.

### 4.1 *Theoretical Analysis of the Algorithms*

Experiments were performed to test the three classifiers on a suite of five datasets normally used in machine learning to compare the performance of algorithms, and maintained by the University of California at Irvine (UCI). The results were compared in terms of predictive accuracy (PA), which indicates the number of correct classifications over the total number of classifications, and relative absolute error (RAE), which is defined as

$$\mathrm{RAE} = \sum_{i=0}^{N} \frac{|P_i - T_i|}{|\bar{T} - T_i|}; \qquad \bar{T} = \sum_{j=1}^{N} T_j, \tag{2}$$

where $P$ is the predicted values, $T$ is the target values. RAE ranges from 0 to infinity, and the best result is obtained when the numerator is equal to 0, leading to $\mathrm{RAE} = 0$.

PA, expressed in %, is usually a good indication of the performance of the algorithm, but does not quantify well the results of very skewed cases. For example, suppose a simple problem with 1000 elements equally distributed among two classes, A and B. If an algorithm classifies all events as class A, it will have a PA of 50%, since half of the classifications are correct. However, its RAE is 100%, as it failed to provide any discriminatory information between the classes. Conversely, if all 1000 events are correctly classified, the PA is equal to 100% and the RAE is equal to 0%.

The five datasets were chosen to test the three algorithms under different conditions, such as number of variables, their domain size, the number of events, and the size of the output classes.

The tests were performed using a 10-fold cross-correlation, where the events are divided into 10 sets of unique events, and each time one of the set is used for learning, and the other 9 used for testing. The process is repeated 10 times, and the results show the average performance of the runs. This cross-correlation testing scheme is used to minimize a bias which might be introduced by selecting specific training and testing events.

Table 3 shows the cumulative results for all three classifiers, along with general information on the size and characteristics of the data. As expected, there is not a clear winner, and all three algorithms are able to outperform the others in specific cases, although very narrowly in most cases. One notable exception is the supermarket dataset, where JRip outperformed by a great margin both NB and J48. In comparison, both NB and J48 have an error rate of 100% and classify all events as members of one class. The main characteristic of this dataset with respect to the others is the larger number of events (4627). JRip also outperforms all other methods using the Glass datasets, which has the characteristics of a larger output class size (6). These results of JRip are particularly important because in the problem of pixel classification, it is usually required to analyze a large number of events, and with a multi-variate output class size.

Although the experimental analysis provides insights on the performance of the algorithms, it is necessary to go beyond PA and RAE and understand what are the main limitations associated with different learning classifiers. Although a full analysis of the learning strategies goes beyond the scope of this paper, a short discussion is provided to direct the reader to the important characteristics that should be taken into consideration.

### 4.1.1 *Learning bias*

Perhaps the most important aspect to take into consideration is the learning or inductive bias. This consists of the learning strategy and set of assumptions that the algorithm uses when learning descriptions.[40] The algorithms used have very different learning biases, which help explain why their performance over the experimental dataset vary. It is important to remember that all classification algorithms used (and most classification algorithms in general) use forms of inductive learning when generating general descriptions from limited observations. Unlike deductive learning, where given the truth values of the premises it is possible to determine with absolute certainty the truth value of the conclusion (modus ponens in logic), in inductive learning the truth value of the learning process cannot be guaranteed. This is why it is very important to understand and test the learned descriptions to assess their validity.

### 4.1.2 *Representation language*

Another very important characteristic is the language used to represent the learned descriptions. Decision rule classifiers, like RIPPER[38] or AQ,[44,45] use a very complex representation language, where descriptions are conjunctions of attribute-value relations. Unlike neural networks, which are black boxes and use a representation language which cannot be easily visualized, decision

**Table 3** Predictive accuracy (PA) (%) and relative absolute error (RAE) (%) computed using a 10-fold cross validation method for the JRip, J48, and NB classifiers over several testing datasets.

| | | | | JRip | | J48 | | NB | |
|---|---|---|---|---|---|---|---|---|---|
| Dataset | # Events PA | # Variables RAE | # Classes PA | RAE | PA | RAE | PA | RAE | PA |
| Ionosphere | 351 | 35 | 2 | 90 | 31 | **91** | **20** | 83 | 38 |
| Glass | 214 | 10 | 6 | **69** | **54** | 67 | 48 | 49 | 73 |
| Diabetes | 768 | 9 | 2 | 76 | 75 | 74 | 69 | **76** | **63** |
| Labor | 57 | 17 | 2 | 77 | 49 | 74 | 70 | **89** | **23** |
| Supermarket | 4627 | 217 | 2 | **77** | **68** | 64 | 100 | 64 | 100 |

Numbers in bold identify the best results.

rules can be inspected and validated by human experts. Although decision tree classifiers, such as C4.5, can convert the learned trees into rules, the resulting descriptions are less compact and composed of atomic relations (e.g., $[x = a]$) and usually cannot include internal disjunctions (e.g., $[x = a \text{ or } b]$).

There is usually a tradeoff between richness of the representation language, ability to represent complex descriptions, and complexity of the algorithms. In general, decision rule algorithms use logical inference to learn descriptions, which is arguably one of the most difficult strategies. However, they are able to learn "parallel-axis descriptions." For example, assuming a Cartesian coordinate system, and points distributed in a circle, a decision rule can only approximate the concept by using rectangles. A transformation of the coordinate system (e.g., into polar coordinates) must be first performed to correctly represent the concept. On the other hand, a neural network is likely to correctly learn a circular pattern, at the expense of not allowing for an inspection of the learned knowledge.

### 4.1.3 *Speed and scalability*

Even with modern computers, the speed at which the algorithms can learn descriptions and their ability to work with ever larger amounts of data is paramount. Mathematical-based methodologies, such as support vector machines (SVM) and Bayesian classifiers, are usually among the fastest methods available because mathematical operations have a low complexity and can be very efficiently implemented. Decision rule classifiers, on the other hand, have a much higher complexity due to the iterative nature of the logic operations.

Scalability depends not only on the number of training events, which is usually one of the most important parameters. In remote sensing pixel classification, in fact, there could be hundreds of thousands or millions of pixels in a single image. However, the number of observations is only one of the dimensions that tests the scalability of the algorithms. The number of attributes in each observation, their type, and domain size also play a crucial role. Some algorithms are in fact better suited to work with numerical values, others work best with categorical ones, and others can work with both types. Some algorithms have been designed specifically to work with attributes with very large domain sets (e.g., attribute $X$ is "items in a supermarket").

### 4.1.4 *Incremental learning*

Decision rule learners have the intrinsic advantage of being able to refine previous rules as new training data become available. Refinement of rules involves adding or dropping conditions in previously learned rules, or splitting a rule in a number of partially subsumed rules. The main advantage is that modification in a rule of the cover for a class does not affect the other rules in the cover. In contrast, although possible, it is more complicated to update a tree as it often involves several updates that propagate from the leafs of the tree all the way to the root. Additionally, the resulting tree might be suboptimal and very unbalanced, prompting for a complete reevaluation of each node. Mathematical-based techniques depend on the strategy used. Methods based on Bayesian inference usually do not cope well, since the *a priori* probabilities must be recomputed as new data becomes available.

### 4.1.5 *Input background knowledge*

One area of study that has received less attention is how to convey to the algorithm a set of conditions or constraints that are known and should be taken into consideration during the learning process. This feature is particularly important when there is an existing knowledge of the data, or constraints on the attributes, which can lead to a simpler knowledge and a faster execution.

## 4.2 *Combining RADARSAT and ASTER Data*

Experiments were performed to test the hypothesis that a combination of RADAR and optical data lead to higher PA than either of the two datasets alone. Different datasets were created to test

the PA of each source individually (A09, R08, and R09), and their various combinations (R08 + A09 and R09 + A09).

Figure 2 shows the PA obtained by different combinations of RADARSAT and ASTER data. The highest PA is found by a combination of RADAR and optical data paired with the JRip and NB classifiers. The PA of the combined RADARSAT and ASTER datasets is about 10% higher than RADARSAT alone and about 20% higher than ASTER alone. There is also a difference between the RADARSAT 2008 and 2009 datasets when used alone, with the newer, wet season data yielding a PA 5% higher than the older, dry season dataset. However, when paired with the ASTER data, the 2009 RADARSAT data are only marginally better. J48 was consistently the lowest performer with any of the datasets, with the exception of the R08 where it outperformed NB, and the only classifier not able to reach 95% accuracy (indicated in Fig. 2 with a horizontal dashed line).

Table 4 shows a summary of all the experiments performed, in terms of number of correct pixel classifications, the resulting PA, the elapsed time, and the classification confusion matrices. There are a total of 14,912 testing pixels (Table 2), and PA is expressed as a fraction of the number of correct classifications divided by the total number of pixels. The highest PA, in excess of 99%, is obtained by the NB with either combination of RADAR and optical data. Only slightly lower is the PA obtained by JRip on the same combined datasets. In terms of elapsed time, NB is the most efficient, being one order of magnitude faster than the other methods. Whereas both decision trees and decision rules formulate hypotheses through an inductive inference process, which consists of computationally expensive generalization and specialization operations, NB learns descriptions by computing probabilities, a much faster operation. However, the quicker execution time is a tradeoff for descriptiveness and understandability of the learned hypotheses.

The result of each experiment is also expressed with a confusion matrix that shows the number of correct and incorrect classifications for each cover class. The rows of the matrix indicate the true class of the pixels, and the columns the predicted (pred.) class of the pixels. Analyzing a confusion matrix is possible to determine what are the most common errors. When using ASTER data alone, all three classifiers have difficulties distinguishing bare soil pixels from urban pixels. This is indicated by the high number of bare soil pixels (1763 JRip; 1604 J48; 1530 NB) misclassified as urban. This class confusion is a result of the dark toned bare soil and urban areas being spectrally similar as many of the structures are constructed with indigenous material.

When using the 2008 RADARSAT data alone, all three classifiers misclassify a large number of agriculture (1933 JRip; 1894 J48; 1890 NB) as urban. This difficulty most likely is a function of the changing conditions within the fields seasonally and that the 2008 dry season image contains few

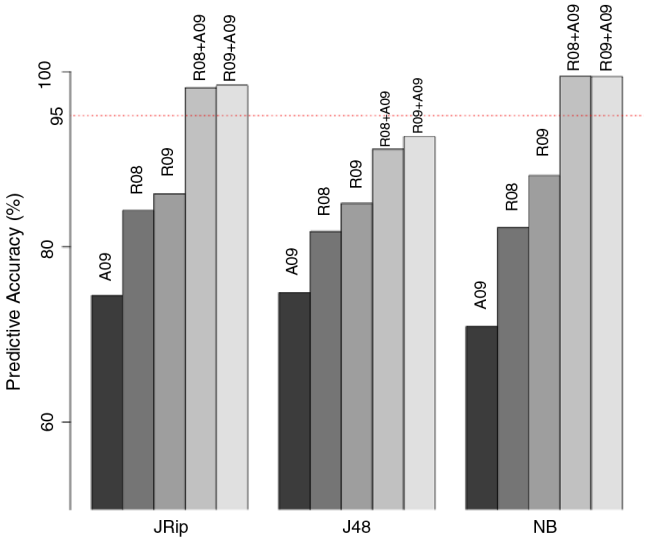



**Fig. 2** Predictive accuracy obtained by the classifiers using different combinations of RADARSAT and ASTER data.

**Table 4** Summary of the experiments performed in terms of number of correct classifications, predictive accuracy (PA), elapsed time and confusion matrix. In the confusion matrix, the columns indicate the predicted (pred.) classifications, and the rows indicate the true classifications. In the data label, R corresponds to RADARSAT and A to ASTER followed by the year of acquisition (e.g., R08 + A09 = RARASAT 2008 plus ASTER 2009).

| | | | | | Pred. | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Dataset | Classifier | Number correct | PA (%) | Time (s) | Agric. | Bare soil | Urban | Water | True |
| R08 | JRip | 11105 | 74.47 | 3.65 | **1941** | **0** | **1933** | **1** | Agriculture |
| | | | | | 0 | **3358** | **33** | **730** | Bare soil |
| | | | | | 775 | 0 | **3138** | **0** | Urban |
| | | | | | 0 | 333 | 2 | **2668** | Water |
| | J48 | 11153 | 74.79 | 1.32 | **1972** | **0** | **1894** | **9** | Agriculture |
| | | | | | 0 | **3453** | **33** | **635** | Bare soil |
| | | | | | 773 | 0 | **3140** | **0** | Urban |
| | | | | | 0 | 413 | 2 | **2588** | Water |
| | NB | 10575 | 70.91 | 0.11 | **2015** | **0** | **1890** | **0** | Agriculture |
| | | | | | 3 | **2849** | **1** | **1268** | Bare soil |
| | | | | | 746 | 0 | **3167** | **0** | Urban |
| | | | | | 0 | 459 | 0 | **2544** | Water |
| R09 | JRip | 12555 | 84.19 | 2.47 | **3647** | **24** | **204** | **0** | Agriculture |
| | | | | | 0 | **2712** | **11** | **1398** | Bare soil |
| | | | | | 456 | 0 | **3457** | **0** | Urban |
| | | | | | 0 | 229 | 35 | **2739** | Water |
| | J48 | 12191 | 81.75 | 0.22 | **3481** | **100** | **271** | **23** | Agriculture |
| | | | | | 0 | **2386** | **239** | **1496** | Bare soil |
| | | | | | 287 | 0 | **3626** | **0** | Urban |
| | | | | | 0 | 262 | 43 | **2698** | Water |
| | NB | 12260 | 82.21 | 0.02 | **3719** | **27** | **129** | **0** | Agriculture |
| | | | | | 70 | **2279** | **1** | **1771** | Bare soil |
| | | | | | 513 | 0 | **3400** | **0** | Urban |
| | | | | | 10 | 131 | 0 | **2862** | Water |
| A09 | JRip | 12836 | 86.07 | 1.3 | **3809** | **0** | **0** | **66** | Agriculture |
| | | | | | 14 | **2304** | **1763** | **40** | Bare soil |
| | | | | | 27 | 164 | **3722** | **0** | Urban |
| | | | | | 0 | 2 | 0 | **3001** | Water |
| | J48 | 12674 | 84.99 | 0.07 | **3868** | **0** | **7** | **0** | Agriculture |
| | | | | | 9 | **2138** | **1604** | **370** | Bare soil |

**Table 4** (*Continued*).

| Dataset | Classifier | Number correct | PA (%) | Time (s) | Pred. Agric. | Pred. Bare soil | Pred. Urban | Pred. Water | True |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | 41 | 207 | **3655** | **0** | Urban |
| | | | | | 0 | 0 | 0 0 | **3003** | Water |
| | NB | 13140 | 88.17 | 0.01 | **3856** | **18** | **1** | **0** | Agriculture |
| | | | | | 0 | **2591** | **1530** | **0** | Bare soil |
| | | | | | 3 | 217 | **3693** | **0** | Urban |
| | | | | | 3 | 0 | 0 | **3000** | Water |
| R08 + A09 | JRip | 14640 | 98.17 | 0.8 | **3696** | **0** | **145** | **34** | Agriculture |
| | | | | | 3 | **4040** | **38** | **40** | Bare soil |
| | | | | | 10 | 0 | **3903** | **0** | Urban |
| | | | | | 0 | 2 | 0 | **3001** | Water |
| | J48 | 13597 | 91.18 | 0.51 | **2650** | **0** | **1225** | **0** | Agriculture |
| | | | | | 3 | **4046** | **32** | **40** | Bare soil |
| | | | | | 15 | 0 | **3898** | **0** | Urban |
| | | | | | 0 | 0 | 0 | **3003** | Water |
| | NB | 14841 | 99.52 | 0.08 | **3817** | **0** | **58** | **0** | Agriculture |
| | | | | | 0 | **4118** | **3** | **1771** | Bare soil |
| | | | | | 6 | 0 | **3907** | **0** | Urban |
| | | | | | 0 | 4 | 0 | **2999** | Water |
| R09 + A09 | JRip | 14688 | 98.49 | 0.58 | **3788** | **16** | **5** | **66** | Agriculture |
| | | | | | 0 | **4012** | **69** | **40** | Bare soil |
| | | | | | 24 | 2 | **3887** | **0** | Urban |
| | | | | | 2 | 0 | 0 | **3001** | Water |
| | J48 | 13813 | 92.63 | 0.16 | **3629** | **0** | **0** | **246** | Agriculture |
| | | | | | 0 | **3304** | **384** | **433** | Bare soil |
| | | | | | 28 | 8 | **3877** | **0** | Urban |
| | | | | | 0 | 0 | 0 | **3003** | Water |
| | NB | 14833 | 99.47 | 0.02 | **3873** | **2** | **0** | **0** | Agriculture |
| | | | | | 35 | **4069** | **17** | **0** | Bare soil |
| | | | | | 4 | 0 | **3909** | **0** | Urban |
| | | | | | 6 | 15 | 0 | **2982** | Water |

Numbers in bold identify the best results.

agricultural fields. Additionally, NB misclassifies several bare soil pixels (1268) as water, while JRip and J48 misclassify about half, 730 and 635, respectively. When using the 2009 RADARSAT data alone, all three classifiers make fewer errors between agriculture and urban, but a larger number of bare soil pixels (1398 JRip; 1496 J48; 1771 NB) are misclassified as water.

When combining RADARSAT and ASTER data, the overall results are significantly better with much fewer classification errors. When using the 2008 RADARSAT and ASTER data, the only large errors are made by J48, classifying 1225 agriculture pixels as urban, and by NB which classifies 1771 bare soil pixels as water. When using the 2009 RADARSAT and ASTER data, JRip and NB obtain nearly perfect PA, whereas J48 misclassifies about 7% of the pixels.

### 4.3 *Band Analysis*

One of the main advantages of using a decision rule algorithm such as JRip is the possibility of inspecting and verifying the learned hypotheses. Equation (3) shows the main rules learned by the JRip classifier using the training set provided.

$$\begin{aligned}
&\text{Water} \leftarrow (\text{Aster.b11} <= 72):\ p = 1431, \quad n = 0 \quad \text{BareSoil} \leftarrow (\text{Aster.b9} <= 78) \quad \text{and} \\
&\quad (\text{RadSatJun09.text.b19} <= 0):\ p = 925, \quad n = 0 \quad \leftarrow (\text{Aster.b10} >= 113) \quad \text{and} \\
&\quad (\text{RadSatJun09.text.b16} <= 17):\ p = 550, \quad n = 0 \\
&\quad \text{Agriculture} \leftarrow \text{Aster.b9} <= 64):\ p = 1569, \quad n = 0 \quad \leftarrow (\text{Aster.b9} <= 72) \quad \text{and} \\
&\quad (\text{RadSatJun09.text.b16} <= 44) \quad \text{and} \quad (\text{RadSatJun09.text.b18} <= 4):\ p = 220, \quad n = 0 \\
&\quad \text{Urban} \leftarrow \text{Default}:\ p = 2731, \quad n = 0.
\end{aligned} \tag{3}$$

JRip generated a total of 14 rules to cover all four classes. The strongest patterns identified, defined as rules that cover at least 50 pixels in the positive class, are shown in Eq. (3). A description of how to read and interpret JRip rules is given in Eq. (1).

In JRip, the classification of an unseen pixel (predicting phase) occurs sequentially. Using the rules in Eq. (3), a pixel is first compared against the rules for the water class, then bare soil and agriculture, and if none of the rules is matched, then the default class is assigned (urban). For example, the first rule for bare soil and the first rule for agriculture both include a condition restricting Aster.b9 band less than 78 and 64, respectively. However, the rules do not overlap, because in the rule for bare soil, which is matched first, there is a conjunction of two conditions, both of which must hold. Therefore, when the rule for agriculture is matched, it is implied that not only Aster.b9 must be less or equal to 64 but RADARSATJun09.text.b19 must also be different than 0.

A very strong pattern was obtained for the water class where a single rule covers 1431 out of 1434 training pixels (Table 2). Except for the first rule (water class), rulesets are made of multiple rules, and in this case it is enough that either of the two rules is satisfied for the classification of the pixel. For both bare soil and agriculture, all rules are made by the conjunction of two or more conditions. This indicates that the combination of two or more bands is necessary to correctly predict a class. In particular, the combinations in the rules include bands from both RADARSAT and ASTER, further indicating that the union of the two datasets is necessary for better results. Finally, the default covers all 2731 examples of the urban class, and no examples of the other classes.

Figure 3 shows the spectral characteristics for the four cover classes in the six bands cited in the JRip rules. No class can be discriminated from the others using a single band, with the exception of the water class that has uniquely low values in the Aster.b11 band. This corresponds to the rules learned by JRip; in fact, the only class covered with a single conjunction is the water class. An inspection of the spectral characteristics of Aster.b11 shows that no class other than water has any pixels with a value less than 72, exactly what was learned by JRip.

### 4.4 *Assessment of RADAR Data in Dry and Wet Season*

Another important aspect of this research is to compare the quality of RADAR data during the dry and wet season. In most experiments performed, the RADAR data for June 2009,

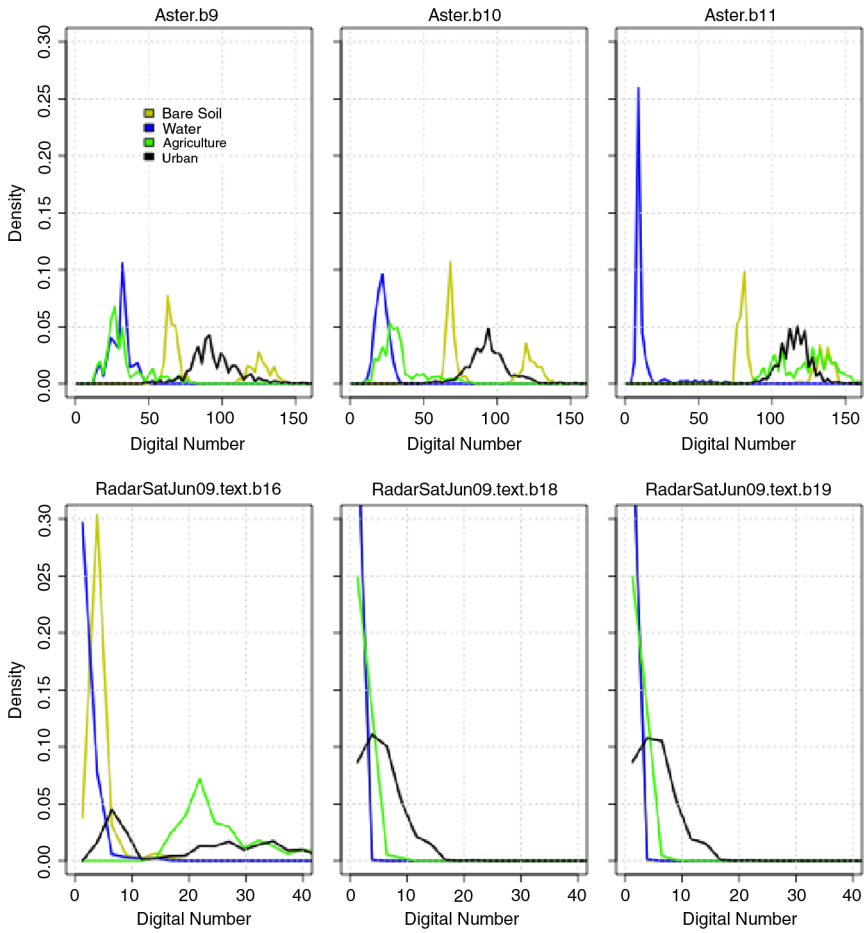



**Fig. 3** Spectral characteristics for the bands used by JRip. Notice the different horizontal scale for the ASTER (top) and RADARSAT (bottom) graphs.

corresponding to the wet season, led to better results, both when used alone and when combined with the ASTER data (see Fig. 2). Significant differences between the December 2008 and June 2009 data are found in the texture bands. In particular, RadSatJun09.text.b5 (December 2008) and the corresponding RadSatJun09.text.b16 (June 2009) differ the most. Because this particular band is instrumental in the classification of agriculture and bare soil pixels, it can explain the large number of bare soil pixels classified as agriculture (158 JRip; 1225 J48; 58 NB) when using R08 + A09, and the very small number of misclassified pixels (5 JRip; 0 J48; 0 NB) when using R09 + A09, shown in Table 4.

The crops during the wet season should be better developed and thus exhibit more texture separating them from the bare soil. During the dry season, those same fields may be fallow and have similar texture response to the bare soil.

The results for the wet season RADAR, June 2009, are better independently and in combination with the ASTER data than the dry season RADAR, December 2008. These differences, however, should be carefully considered as they are to some degree, and likely a considerable degree, a function of changing conditions in the fields during the year. During December many of the fields have been harvested and are fallow until the following summer. These fallow fields are spectrally very similar and confused with the bare soil as indicated in Table 4.

## 4.5 *Prediction of the Entire Scene*

A sample of 62,500 pixels were selected uniformly over the domain and classified according to the hypotheses learned by the three classifiers. Figure 1(b), 1(c), and 1(d) shows the entire scene classified into the four cover classes by JRip, J48, and NB, respectively. Although all three

classifiers correctly predict the cover of the training and testing areas, JRip seemed to perform best due to its ability to recognize the river very well. NB, on the other hand, is not able to classify correctly the water pixels, and only a very small portion of the river is identified.

## 5 Conclusions

Supervised machine learning classifiers were employed with a fused RADAR and optical dataset to predict pixels into one of four classes, agriculture, bare soil, urban, and water. Three classifiers were employed, based on decision rules, decision trees, and naive Bayesian methodologies. Each classifier was trained and tested over selected regions, and their results compared.

In general, the decision tree rule induction algorithm outperformed the other methods, achieving both a very high predictive accuracy and overall good entire scene prediction. This is an important result because decision rules have historically received less attention than the other classifier methods.

All three classifiers performed best with combined RADAR and optical data. This result is consistent with results published in the literature that point out that radar information, especially texture, can lead to better land cover classifications.

## Acknowledgments

The RADARSAT-2 imagery acquisition was funded by a research grant submitted to the Canadian Space Agency under the Science and Operational Applications Research (SOAR) program, ID number 3126. Work performed under this project has been partially supported by George Mason University Summer Research Funding. A special thanks to Savika Voratanitkitkul for her comments and suggestions, and for proofreading the manuscript.

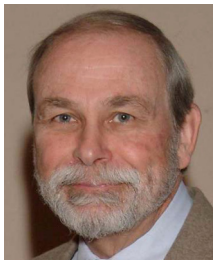
## References

1. J. Rogan et al., "Mapping land-cover modifications over large areas: a comparison of machine learning algorithms," *Remote Sens. Environ.* **112**(5), 2272–2283 (2008), http://dx.doi.org/10.1016/j.rse.2007.10.004.
2. M. Hall et al., "The WEKA data mining software: an update," *SIGKDD Explor.* **11**(1), 10–18 (2009), http://dx.doi.org/10.1145/1656274.
3. J. A. Benediktsson, P. H. Swain, and O. K. Ersoy, "Neural network approaches versus statistical methods in classification of multisource remote sensing data," *IEEE Trans. Geosci. Remote Sens.* **28**(4), 540–552 (1990), http://dx.doi.org/10.1109/TGRS.1990.572944.
4. S. B. Serpico and F. Roli, "Classification of multisensor remote-sensing images by structured neural networks," *IEEE Trans. Geosci. Remote Sens.* **33**(3), 562–578 (1995), http://dx.doi.org/10.1109/36.387573.
5. P. Bosdogianni, M. Petrou, and J. V. Kittler, "Mixed pixel classification with robust statistics," *IEEE Trans. Geosci. Remote Sens.* **35**(3), 551–559 (1997), http://dx.doi.org/10.1109/36.581966.
6. M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Decision fusion for the classification of urban remote sensing images," *IEEE Trans. Geosci. Remote Sens.* **44**(10), 2828–2838 (2006), http://dx.doi.org/10.1109/TGRS.2006.876708.
7. S. Aksoy et al., "Land cover classification with multi-sensor fusion of partly missing data," *Photogramm. Eng. Remote Sens.* **75**(5), 577–593 (2009).
8. V. Walter, "Object-based classification of remote sensing data for change detection," *ISPRS J. Photogramm. Remote Sens.* **58**(3–4), 225–238 (2004), http://dx.doi.org/10.1016/j.isprsjprs.2003.09.007.
9. C. Q. Zhu and X. M. Yang, "Study of remote-sensing image texture analysis and classification using wavelet," *Int. J. Remote Sens.* **19**(16), 3197–3203 (1998), http://dx.doi.org/10.1080/014311698214262.

10. S. Ryherd and C. Woodcock, "Combining spectral and texture data in the segmentation of remotely sensed images," *Photogramm. Eng. Remote Sens.* **62**(2), 181–194 (1996).
11. R. Gaetano, G. Scarpa, and G. Poggi, "Hierarchical texture-based segmentation of multiresolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.* **47**(7), 2129–2141 (2009), http://dx.doi.org/10.1109/TGRS.2008.2010708.
12. D. Puig and M. García, "Pixel-based texture classification by integration of multiple texture feature evaluation windows," *Lec. Notes Comput. Sci.* **2652**, 793–801 (2003), http://dx.doi.org/10.1007/b12122.
13. F. Henderson et al., "Evaluation of SAR-optical imagery synthesis techniques in a complex coastal ecosystem," *Photogramm. Eng. Remote Sens.* **68**(8), 839–846 (2002).
14. J. Toyra, A. Pietroniro, and J. Martz, "Multisensor hydrologic assessment of a freshwater wetland," *Remote Sens. Environ.* **75**(2), 162–173 (2001), http://dx.doi.org/10.1016/S0034-4257(00)00164-4.
15. F. Dell'Acqua, P. Gamba, and G. Lisini, "Improvements to urban area characterization using multitemporal and multiangle SAR images," *IEEE Trans. Geosci. Remote Sens.* **41**(1), 153–159 (2003), http://dx.doi.org/10.1109/TGRS.2002.807754.
16. A. Banner and F. Ahern, "Incident angle effects on the interpretability of forest clearcuts using airborne C-HH SAR imagery," *Can. J. Remote Sens.* **2**(1), 64–66 (1995).
17. S. Hegarat-Mascle et al., "Application of Shannon information theory to a comparison between L- and C-band SIR polarimetric data versus incident angle," *Remote Sens. Environ.* **60**(2), 121–130 (1997), http://dx.doi.org/10.1016/S0034-4257(96)00164-2.
18. Y. Gauthier, M. Bernier, and J. Fortin, "Aspect and incident angle sensitivity in ERS-1 SAR data," *Int. J. Remote Sens.* **19**(10), 2001–2006 (1998), http://dx.doi.org/10.1080/014311698215117.
19. J. Campbell, *Introduction to Remote Sensing*, 4th ed., Guilford Publications, New York (2007).
20. H. Anys and D. He, "Evaluation of textural and multipolarization radar features for crop classification," *IEEE Trans. Geosci. Remote Sens.* **33**(5), 1170–1181 (1995), http://dx.doi.org/10.1109/36.469481.
21. T. Kurosu et al., "Texture statistics for classification of land use with multitemporal JERS-1 SAR single-look imagery," *IEEE Trans. Geosci. Remote Sens.* **37**(1), 227–235 (1999), http://dx.doi.org/10.1109/36.739157.
22. A. H. S. Solberg and A. K. Jain, "Texture fusion and feature selection applied to SAR imagery," *IEEE Trans. Geosci. Remote Sens.* **35**(2), 475–479 (1997), http://dx.doi.org/10.1109/36.563288.
23. P. Townsend, "Estimating forest structure in wetlands using multitemporal SAR," *Remote Sens. Environ.* **79**(2–3), 288–304 (2002), http://dx.doi.org/10.1016/S0034-4257(01)00280-2.
24. B. Haack et al., "Radar and optical data comparison/integration for urban delineation: a case study," *Photogramm. Eng. Remote Sens.* **68**(12), 1289–1296 (2002).
25. N. D. Herold, B. N. Haack, and E. Solomon, "Radar spatial considerations for land cover extraction," *Int. J. Remote Sens.* **26**(7), 1383–1401 (2005), http://dx.doi.org/10.1080/01431160512331337998.
26. A. N. Nyoungui, E. Tonye, and A. Akono, "Evaluation of speckle filtering and texture analysis methods for land cover classification from SAR images," *Int. J. Remote Sens.* **23**(9), 1895–1925 (2002), http://dx.doi.org/10.1080/01431160110036157.
27. H. Huang, J. Legarsky, and J. Othman, "Land-cover classification using Radarsat and Landsat imagery for St. Louis, Missouri," *Photogramm. Eng. Remote Sens.* **73**(1), 37–43 (2007).
28. L. Zhang et al., "Texture feature fusion with neighborhood oscillating tabular search for high resolution image classification," *Photogramm. Eng. Remote Sens.* **74**(1), 323–331 (2008).
29. D. Leckie, "Synergism of synthetic aperture radar and visible/infrared data for forest type discrimination," *Photogramm. Eng. Remote Sens.* **56**(9), 1237–1246 (1990).
30. S. Pal, T. Majumdar, and A. Bhattachrya, "ERS-2, SAR and IRS-IC LISS III data fusion: a PCA approach to improve remote sensing based geological interpretation," *J. Photogramm. Remote Sens.* **61**(5), 281–297 (2007), http://dx.doi.org/10.1016/j.isprsjprs.2006.10.001.

31. C. Santos and J. Messina, "Multi-sensor data fusion for modeling African palm in the Ecuadorian Amazon," *Photogramm. Eng. Remote Sens.* **74**(6), 711–723 (2008).
32. A.H.S. Solberg, A. K. Jain, and T. Taxt, "Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images," *IEEE Trans. Geosci. Remote Sens.* **32**(4), 768–778 (1994), http://dx.doi.org/10.1109/36.298006.
33. A. Saraf, "IRS-IC-LISS-III and PAN data fusion: an approach to improve remote sensing based mapping techniques," *Int. J. Remote Sens.* **20**(10), 1929–1934 (1999), http://dx.doi.org/10.1080/014311699212272.
34. G. Simone et al., "Image fusion techniques for remote sensing application," *Information Fusion* **3**(1), 3–15 (2002), http://dx.doi.org/10.1016/S1566-2535(01)00056-2
35. I. H. Witten et al., "Weka: practical machine learning tools and techniques with java implementations," in *Proc. ICONIP/ANZIIS/ANNES'99 Workshop Emerg. Knowl. Eng. and Connectionist-Based Inform. Syst.*, Hamilton, New Zealand, N. Kasabov and K. Ko, Eds., pp. 192–196 (1999).
36. E. Frank et al., "WEKA—a machine learning workbench for data mining," in *The Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds., pp. 1305–1314, Springer, New York, NY (2005).
37. I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed., Morgan Kaufmann, San Francisco (2005).
38. W. W. Cohen, "Fast effective rule induction," in *Proc. 12th Intl. Conf. Mach. Learn.*, Orlando, FL, A. Prieditis and S. Russell, Eds., pp. 115–123 (1995).
39. J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA (1993).
40. T. Mitchell, *Machine Learning*, McGraw Hill, Boston, MA (1997).
41. J. R. Quinlan, "Improved use of continuous attributes in C4.5," *J. Artif. Intell. Res.* **4**, 77–90 (1996).
42. S. Russell and P. Norvig, *Artificial Intelligence, A Modern Approach*, Prentice Hall, Upper Saddle River, New Jersey (1995).
43. A. Gelman et al., *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton (2003).
44. G. Cervone, L. Panait, and R. Michalski, "The development of the AQ20 learning system and initial experiments," in *Proc. of the Fifth International Symposium on Intelligent Information Systems*, Zakopane, Poland, p. 13 (2001).
45. G. Cervone, P. Franzese, and A. P. Keesee, "Algorithm quasi-optimal (AQ) learning," *WIREs: Comput. Stat.* **2**(2), 218–236 (2010), http://dx.doi.org/10.1002/wics.78.

**Guido Cervone** is associate professor of Geoinformatics at George Mason University in Fairfax, Virginia. His main interest is the spatial and temporal data mining of large remote sensing and model data. He worked in the theoretical development and implementation of symbolic and evolutionary machine learning systems, and their application for natural hazards monitoring, analysis and forecasting. He has been a visiting scientist at the National Center for Atmospheric Research, in Boulder, CO, and at the Goethe Universitat Frankfurt, Germany.



**Barry Haack** is a professor of Geographic and Cartographic Sciences at George Mason University in Fairfax, Virginia. He was educated in Geography at the University of Wisconsin, San Diego State University and the University of Michigan. He has been a visiting scientist at the United State Geologic Survey and has had fellowships at NASA Goddard, the U. S. Air Force, and the Jet Propulsion Laboratory at the California Institute of Technology. He was a scientist at the Regional Remote Sensing Facility in Kenya, a Fulbright professor at the University of Dar es Salaam in Tanzania and a visiting scientist at the International Centre for Integrated Mountain Development in Nepal. He has served as a consultant for various UN agencies, the World Bank, USAID, and several country governments as well as conducted workshops in Afghanistan, Viet Nam, Nepal, Syria, and Swaziland.