Developing Collaborative Classifiers using an Expert-based Model

Giorgos Mountrakis, Raymond Watts, Lori Luo, and Jida Wang

Abstract

This paper presents a hierarchical, multi-stage adaptive strategy for image classification. We iteratively apply various classification methods (e.g., decision trees, neural networks), identify regions of parametric and geographic space where accuracy is low, and in these regions, test and apply alternate methods repeating the process until the entire image is classified. Currently, classifiers are evaluated through human input using an expert-based system; therefore, this paper acts as the proof of concept for collaborative classifiers. Because we decompose the problem into smaller, more manageable sub-tasks, our classification exhibits increased flexibility compared to existing methods since classification methods are tailored to the idiosyncrasies of specific regions. A major benefit of our approach is its scalability and collaborative support since selected low-accuracy classifiers can be easily replaced with others without affecting classification accuracy in high accuracy areas. At each stage, we develop spatially explicit accuracy metrics that provide straightforward assessment of results by non-experts and point to areas that need algorithmic improvement or ancillary data. Our approach is demonstrated in the task of detecting impervious surface areas, an important indicator for human-induced alterations to the environment, using a 2001 Landsat scene from Las Vegas, Nevada.

Introduction

Current image classification methods provide choices among many complex algorithms. At one extreme, algorithms provide great internal sophistication and little opportunity for human involvement in classification logic; at the other extreme, humans can examine and interpret each classification step. Neural networks are an example of the former type and decision trees an example of the latter. Algorithms that offer little opportunity for human involvement have been observed to classify accurately and efficiently in selected areas, but are not currently selected for large-scale implementations; for example, the National Land Cover Dataset (NLCD) is based on an inductive machine learning decision tree approach. We conclude that there is a poorly understood interplay between image complexity, algorithmic complexity, and classification error that offers opportunities for synthesis of improved performance through application of different algorithms to different regions of geographic and parameter space. We introduce a strategy for applying such a hybrid process.

The central premises of this paper are: (a) that complexity is not high in all parts of the input space, (b) that different algorithms can be applied in different parts, thereby adaptively matching algorithmic complexity to image complexity, and (c) that classification accuracy improvements can be achieved with this approach by establishing a framework for progressive accuracy increase using hybrid classifiers. These premises translate to three simple operating principles that are admittedly cumbersome in application at this early stage of evolution:

1. Perform diagnostics that identify optimum methods to use from the constantly evolving suite of "standard" methods.
2. As part of these diagnostics, consider alternative inputs.
3. Apply the diagnostics iteratively to parts of the input space where earlier processing stages did not work well.

Upon automation, the potential for improved performance is great. The worst performance from the hybrid method is the best performance from the single standard methods; this is a strong starting point.

This paper acts as the proof-of-concept for a collaborative framework for integration of multiple classifiers. A collaborative framework divides the problem into multiple sub-problems and addresses the classification in each of the different parts of the input space through evaluation of multiple competing algorithms. We think the concept of a collaborative framework is a novel contribution in the remote sensing field even at this early stage. In this preliminary phase, our approach is implemented in an expert-system environment requiring human involvement. Expert-based systems have already been implemented successfully in remote sensing applications; for an early review in image processing applications, see Matsuyama (1989), and for a good explanation from the remote sensing perspective, see Tso and Mather, (2001) and Jensen (2005). Because our hybrid algorithms use machine-aided human judgment to select classifiers and their inputs, there is considerable opportunity for human involvement and application of specialized knowledge in these selections, therefore adding strength to the evaluation of potential classifiers and inputs.

Giorgos Mountrakis, Lori Luo, and Jida Wang are with the Department of Environmental Resources and Forest Engineering, College of Environmental Science and Forestry, State University of New York, Syracuse, NY (gmountrakis@esf.edu).

Raymond Watts is with the Fort Collins Science Center, United States Geological Survey, Fort Collins, CO.

Photogrammetric Engineering & Remote Sensing
Vol. 75, No. 7, July 2009, pp. 831–843.

0099-1112/09/7507–0831/$3.00/0

One of the premises of a collaborative framework is the variability of classification complexity (Mountrakis, 2008). This is illustrated by the relative ease of identifying water and some types of vegetation and the relative difficulty of separating other land types (e.g., soil from constructed areas). In this paper, we will demonstrate our collaborative classification approach on the task of constructed impervious surface identification. Our methodology can easily be generalized to classify other remotely-sensed datasets and has the potential to find fruitful ground in non-imagery classification problems. We should note that significant improvements in classification accuracy are not the major target of this work, rather we focus on introducing a general framework with at least as accurate results: a standard that is practically guaranteed, as noted above, because the worst performance of hybrid methods should be as good as the best performance of single standard methods.

We focus on producing binary maps of manmade impervious surface areas (ISAs): roads, sidewalks, parking lots, driveways, and rooftops. ISAs introduce substantial functional changes in natural environments and ecosystems; they reduce or eliminate the capacity of underlying soil to absorb water, increase peak discharges associated with storm and snowmelt events, increase likelihood of downstream flooding (Berry and Horton, 1974), and transport pollutants that degrade biological, chemical, and physical characteristics of lakes, streams, and estuaries, particularly downstream from urban areas. Biological integrity of streams and riparian habitats diminishes as the fractional ISA surrounding them increases (Kennen, 1999). Roads, whose surfaces are an important ISA, provide the skeleton for construction of other types of ISAs. Roads themselves have been shown to produce dozens of adverse ecological changes (Forman, 2003; Trombulak and Frissell, 2000; Forman and Deblinger, 2000; Forman and Alexander, 1998; Spellerberg, 2002). Thus, our investigations have significant potential value in the evaluation of human alteration of natural environments.

The balance of this paper presents current remote sensing classification methods (in the next Section), followed by a description of the study area and preliminary data processing. Next, we describe the details of our collaborative classification, and present our results, followed by a discussion of benefits and limitations of our methods.

Existing Appr oaches

This paper discusses an alternative approach for remote sensing classification and is demonstrated on impervious surface area (ISA) detection. Our literature review includes existing methods for imperviousness detection, works on uncertainty incorporation in classification results, and hybrid models where multiple algorithms are employed.

Forster (1980) was one of the first to examine the relationship between Landsat data and percentages of land-cover types. The analysis was performed in the Sydney metropolitan area using multiple regression techniques. Much later, Ridd (1995) evaluated the accuracy of mapping the percentage of ISA using 30 m resolution Thematic Mapper multispectral imagery. With this approach, detailed land-cover, land-use, and biophysical parameters were obtained for urban ecosystems. In the recent years, a plethora of algorithms have surfaced on ISA detection (Ji and Jensen, 1999; Ward *et al.*, 2000; Wang *et al.*, 2000; Flanagan and Civco, 2001; Smith and Goetz, 2001; Small, 2002; Wu and Murray, 2003; Hodgson *et al.*, 2003; Bauer *et al.*, 2005; Yang, 2006; Powell *et al.*, 2007; Aitkenhead and Dyer, 2007). The latest addition is a book devoted exclusively to remote sensing application towards impervious surface identification (Weng, 2007). The aforementioned works have a variable accuracy ranging from 75 percent to 94 percent. Our approach complements these works by adding a spatially-explicit accuracy layer to the obtained results while concurrently exhibiting comparable accuracy.

The only existing methodology in the literature applied at a nationwide scale is the approach currently used to classify the imperviousness in the NLCD dataset (Yang *et al.*, 2003; Homer *et al.*, 2004). The authors who described NLCD processing recognized that their decision-tree approach overestimated ISA in rural areas, and as a result, the NLCD program employs manual editing to eliminate ISA outside of urban areas. There is, of course, ISA in rural areas, and the zero expression of rural ISA in NLCD suggests a need for algorithmic improvement. Other approaches have been implemented to incorporate texture/size/shape analysis, mostly using e-Cognition® software. A comparison of object-based and per-pixel classifiers for ISA detection is available in Yuan and Bauer (2006). Our method does not compete directly with object-based analysis, but actually establishes a framework where these computationally-expensive metrics are implemented only in complex situations.

Recent work has focused on deriving the uncertainty map of the land-cover prediction. These works present uncertainty of land-cover classification for each pixel. Foody *et al.* (1992) and Canters (1997) applied class membership values in the maximum-likelihood classifier to model the classification uncertainty. Steele *et al.* (1998) developed an approach to constructing uncertainty maps based on misclassification probabilities. Carpenter *et al.* (1999) used a voting ARTMAP system to produce uncertainty map for each pixel. Pontius (2000) presented a method to separate overall classification error into quantification error and location error. Alimohammadi *et al.* (2004) used maximum likelihood classification algorithm to perform the classification and generated uncertainty estimation. Liu *et al.* (2004) provided a method on uncertainty of large-scale mapping based on disagreement between different methods. Aires *et al.* (2004) developed a neural network approach for remote sensing classification that incorporates network uncertainty.

Hybrid models which merge multiple classifiers together are adopted to increase classification accuracy. Several works in machine learning have showed the potential of a hybrid approach (Hansen and Salamon, 1990; Perrone, 1992; Wolpert, 1992). A simple implementation of the hybrid concept is to average predictions of different classifiers (Krogh and Vedelsby, 1995; Breiman, 1996). More advanced methods (Steele, 2000) used a product rule, sum rule, and stacked regression methods to optimally merge multiple methods; this hybrid approach delivered higher accuracies (ranging from 66 percent to 71 percent) than individual classifiers (ranging from 46 percent to 60 percent). More recently, Coe *et al.* (2005) developed a hybrid model combining an object-oriented and a pixel-based approach. They fused Landsat, Ikonos, and lidar data and parcel data together to establish a spatial urban object database at multiple spatial scales and class resolutions. Further evaluation of the hybrid methods shows that multiple classifiers may not only behave differently globally, but may also have distinct local behavior (Jain *et al.*, 2000). Motivated by this work, Liu *et al.* (2004) presented a hybrid classification approach using decision tree and ARTMAP neural network, and applied them on North America using AVHRR data. Their work moves beyond the winner-takes-all methodology and into a fuzzy merging of multiple classifier outputs by using three levels of agreement and disagreement. From commercial software, Definiens' Enterprise Image Intelligence Suite supports hierarchical rules and allows application of

different algorithms at different parts of the input space. It has limited modeling capabilities, as only simple nearest neighbor classifiers are included in the software.

There is one downfall of current hybrid models and recent classifiers associated with uncertainty metrics; namely, they both increase significantly the internal algorithmic complexity. Furthermore, these works recognize that for a hybrid method to work, classifiers need to make their errors in different parts of the input space (Krogh and Vedelsby, 1995). However, from the algorithmic design perspective, there are no adjustments to force algorithms to focus on different parts of the problem; instead researchers have employed a post-process analysis to identify where each classifier may outperform another. By doing so, potential benefits are a random result of each classifiers behavior instead of a calculated design. Our goal is to create a classification framework for collaborative contributions of multiple algorithms. What separates our work from others is (a) our ability to incorporate context behind the purpose of each classifier in a hybrid approach (in our application, for example, to separate dark ISA from dark soil), and (b) the progressive adjustment of complexity based on the underlying problem characteristics leading to application of complex algorithms in selected parts of the parametric and geographical space. The major benefit of our method is that algorithms can be seen as plug-ins and be replaced/improved depending on problem requirements; for further discussion see Mountrakis (2008).

Targeted Outcome, Study Ar ea, and Classification Inputs

From the implementation perspective, our goal is to interpret the presence or absence of impervious surface areas (ISA) within 30 m Landsat cells. We selected coarser Landsat imagery as opposed to high-resolution imagery because our method is designed for large-scale applications. Our binary product goal contrasts with other ISA interpretation efforts, including the sub-pixel fractional ISA interpretation produced by NLCD. The rationale for our choice is that the typical 15 percent error rate of sub-pixel ISA determinations concentrates primarily in the lower ISA percentages, i.e., primarily in rural areas. The overestimation of rural ISA was recognized by Yang *et al.*, (2003), who applied ancillary data such as population density and buffered roads, NLCD 1992, and NOAA City Lights to build an urban mask that was used to zero all rural ISA estimates. Considering potential uses of ISA maps and the fact that rural areas cover 90 to 95 percent of the U.S. landscape, we believe that a large-scale binary product potentially offers a more balanced outcome (rural and urban areas) that retains most of the essential spatial pattern information contained in fractional ISA products. For applications interested in a fractional product, our binary product can act as an image-based mask for selected application of subpixel algorithms.

Our study area is approximately a 49 km × 57 km rectangle containing Las Vegas, Nevada and its immediate suburbs (Figure 1). More specifically, we used a subset of an April 2001 Landsat scene (1,905 × 1,644 pixels), and the six bands with approximately 30 m ground pixel resolution.

In addition to the established Tasseled Cap transformations (Brightness, Greenness, and Wetness), and the Normalized Difference Vegetation Index (NDVI), we included the following classification inputs in order to discriminate ISA presence:

1. Normalized differences between other bands (e.g., normalized difference between bands 5 and 1).
2. Multiple band-to-band advanced correlations (e.g., difference between band 1 and absolute Greenness normalized by the summation of bands 1 and 5).
3. Neighborhood metrics using a standard deviation mask (e.g., 5 × 5, 7 × 7) to distinguish ISA from other features with similar spectral response but distinct neighboring characteristics.

We tested 277 different inputs expressing numerous of the aforementioned combinations, and we finally selected 23 for our approach. Table 1 displays the 23 inputs and their calculation method.

An Expert-based Method for Collaborative Algorithms

The major motivation of our work is to establish a framework where multiple algorithms can work collaboratively and replacement of one algorithm does not affect the

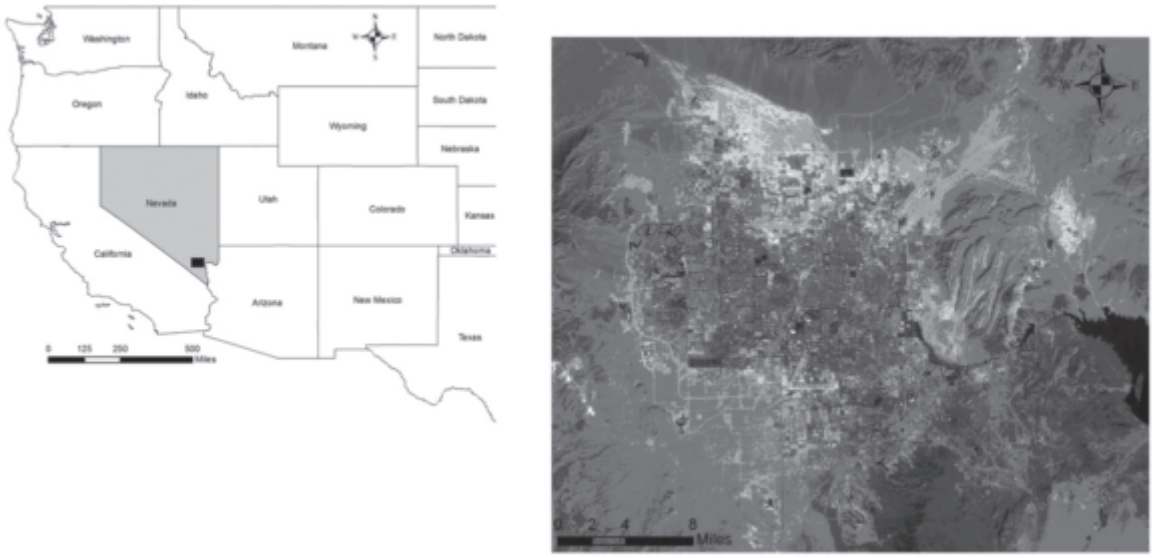

Figure 1. Study area and Landsat Image. A color version of this figure is available at the website: **www.asprs.org**. ASPRS

TABLE 1. CLASSIFICATION INPUT TABLE

Abbreviation	Calculation Method (P1 to P6 are the six Landsat bands)
B	Brightness = (0.3561 *P1) + (0.3972 *P2) + (0.3904 *P3) + (0.6966 *P4) + (0.2286 *P5) + (0.1596 *P6)
G	Greenness = (−0.3344 *P1) + (−0.3544 *P2) + (−0.4556 *P3) + (0.6966 *P4) + (−0.0242 *P5) + (−0.2630 *P6)
ND41	ND41 = (P4 − P1) / (P4 + P1)
ND51	ND51 = (P5 − P1) / (P5 + P1)
ND54	ND54 = (P5 − P4) / (P5 + P4)
ND42	ND42 = (P4 − P2) / (P4 + P2)
NDB5	NDB5 = (Brightness − P5) / (Brightness + P5)
NDaG1	NDaG1 = (abs(Greenness) − P1) / (abs(Greenness) + P1)
NDaG5	NDaG5 = (abs(Greenness) − P5) / (abs(Greenness) + P5)
D51sd5	Local standard deviation on a 5 × 5 moving window as applied on P5 − P1
ND45sd3	Local standard deviation on a 3 × 3 moving window as applied on ND45
ND43sd5	Local standard deviation on a 5 × 5 moving window as applied on ND43
ND43sd7	Local standard deviation on a 7 × 7 moving window as applied on ND43
ND43sd9	Local standard deviation on a 9 × 9 moving window as applied on ND43
ND52sd3	Local standard deviation on a 3 × 3 moving window as applied on ND52
NDB1sd5	Local standard deviation on a 5 × 5 moving window as applied on NDB1
Bsd3	Local standard deviation on a 3 × 3 moving window as applied on Brightness
UBaG5	UBaG5 = (Brightness − abs(Greenness)) / (P5)
UBaG1	UBaG1 = (Brightness − abs(Greenness)) / (P1)
UND1aG15	UND1aG15 = (P1 − abs(Greenness)) / (P1 + P5)
UNtanB4	UNtanB4 = tan(Brightness − P4)
UNtanBWW3	UNtanBWW3 = tan(Brightness − Wetness) − tan(Wetness − P3) where Wetness = (0.2626 *P1) + (0.2141 *P2) + (0.0926 *P3) + (0.0656 *P4) + (−0.7629 *P5) + (−0.5388 *P6)
UNtanBGWB	UNtanBGWB = tan(Brightness − Greenness) − tan(Wetness − Brightness)

Note: Moving window algorithms used symmetrical buffers along image edges to compensate for lack of pixels underneath each window at image borders. The abbreviations of the last four dimensions start with "U" to denote uncommon dimension. Also, "ND" stands for Normalized difference, and "sd" relates to a standard deviation mask. Finally in our calculations "abs" stands for absolute value, and "tan" relates to tangential value.

operation of others. By doing so, limitations of one approach can be easily identified, and incremental improvements can be achieved as classifiers improve. Our method targets large-scale complex classifications, such as the NLCD where diverse landscape characteristics mandate flexible, easily updated, yet accurate methods. Furthermore, an inherent outcome of the application of multiple algorithms is the accuracy variability in the classified product. Spatial error estimation is a highly desired characteristic, because large-scale maps are expected to act as inputs for further analysis by scientists; e.g., urban modelers, biologists, and hydrologists, who are not familiar with image classification methods and their limitations.

In this paper, we demonstrate the collaborative framework concept through the application of impervious surface detection. Our task is to create a binary product expressing impervious surface presence or absence. Considering that our algorithm may act as a filtering process for subsequent subpixel classifiers, we assign in the impervious class pixels with any portion of imperviousness. At the heart of our approach is an expert-based system (EBS) that partitions the classification problem, thus isolating various parts of the input space for application of tailored classification methods. Currently, the partitioning of the problem in several sub-problems is assigned through human interaction, but we envision in the near future an automated process. The EBS starts with simple algorithms (e.g., two-dimensional classifiers) and tests numerous combinations of inputs at low dimensional space. EBS prioritizes these numerous combinations based on classification accuracy and portion of the scene classified. Up to this point, the process is automated. At the next step the results are presented to an expert, who using visualization tools, attempts to assign context behind the absorbed pixels (e.g., water). This is an iterative process where the expert gradually assigns context to classified pixels until the whole scene is classified. As classification complexity increased, we raise the inputs and algorithmic complexity as well. There is a stage in that iterative process that two-dimensional classifiers will not have enough discriminatory power to classify remaining pixels. If that happens, we resort to higher dimensions and more powerful algorithms (e.g., neural networks, decision trees) to help with the classification process. Because of the dimensionality increase, it is difficult for an expert to assign context at the input space. In that case, we assign context based on output performance of the algorithm, for example, how many pixels were classified correctly in low, medium and high brightness.

The expert system segments the input space into regions where different processes will be applied (spectral clustering). The key to the sub-problem identification is to perform it in a context-specific manner. Our definition of context-specific classification relates to algorithms targeting specific tasks; it does not relate to geographical context (e.g., spatial

relationships among features). In our case, we select to gradually distinguish impervious surface areas (ISAs) from vegetation, water, and soil. We have also selected to further distinguish ISAs into low, medium, and high brightness.

The hierarchical structure of the expert system is presented in Figure 2. All pixels are originally inserted into the algorithm, and as they propagate through the hierarchical structure, they get classified. The major differences between our expert-based hierarchical approach and automated inductive machine learning decision trees are:

1. Specific context is assigned to each node. Nodes are designed to perform specific tasks (e.g., separate ISA from vegetation) instead of being identified through a mathematical process that has no such context motivation.
2. Partitioning at each node is not restricted to one-dimensional linear decision (e.g., NDVI >0.1). Multiple dimensions can be used concurrently. These dimensions can represent specific spectral bands, or calculated spectral transformations (e.g., principle components) along with texture information (e.g., a standard deviation mask). In addition, non-linear partitioning algorithms can be implemented within each node.
3. A node on our hierarchical structure does not necessarily assign an output class; it may provide an output (a leaf) for down-tree use by other classifiers (e.g., a neural network may process input pixels from a higher node).
4. A node does not only support dichotomy, instead more than two partitions can be identified from a single node.
5. A classifier at any node level may be any type of algorithm, ranging from simple linear regression to complex neural networks. Automated designs for decision trees only support linear regression at the leaf level.
6. Our hierarchical structure is a semi-automated procedure that utilizes user-provided input to assign pixels, based on all available data at that level of the tree, to processing nodes. Inductive machine learning trees fully automate the training process and incorporate little or no user input.

The variables shown in Table 1 were selected from a much larger array of variables derived from the six input TM bands (mathematically these are not independent variables, but they are treated as independent in our supervised iterative process). There are an infinite number of linear band combinations, band ratios, ratios of linear combinations, neighborhood statistics and their ratios, and so on. We evaluated 277 derived variables by examining them in pairs. The human supervisor thus dealt with individual two-dimensional (2D) representations out of 277 preselected inputs. For each pair of variables we generated 50 rows $\times$ 50 columns 2D histograms of the ISA and non-ISA pixel populations, and then further simplified the 2D picture by displaying the ratio of ISA to non-ISA pixel counts in each histogram cell. Many parameter pairs exhibited poor separation of high- and low-ISA regions; our objective was to find the few 2D parameter pairs that showed clear clustering, and then to express the dividing line(s) between the clusters mathematically. Figure 3 shows a parameter pair that exhibits good clustering and the mathematical functions that separated the zones of clustering. Many parameter pairs displayed such poor clustering that the pair was simply rejected. For promising pairs, the mathematical separation rules were applied to individual pixels, the accuracy of the separation calculated, and this accuracy was used as a figure of merit for the parameter pair. We were able to reject parameter pairs that were less than 90 percent

Figure 2. Hierarchical context-specific Classification.

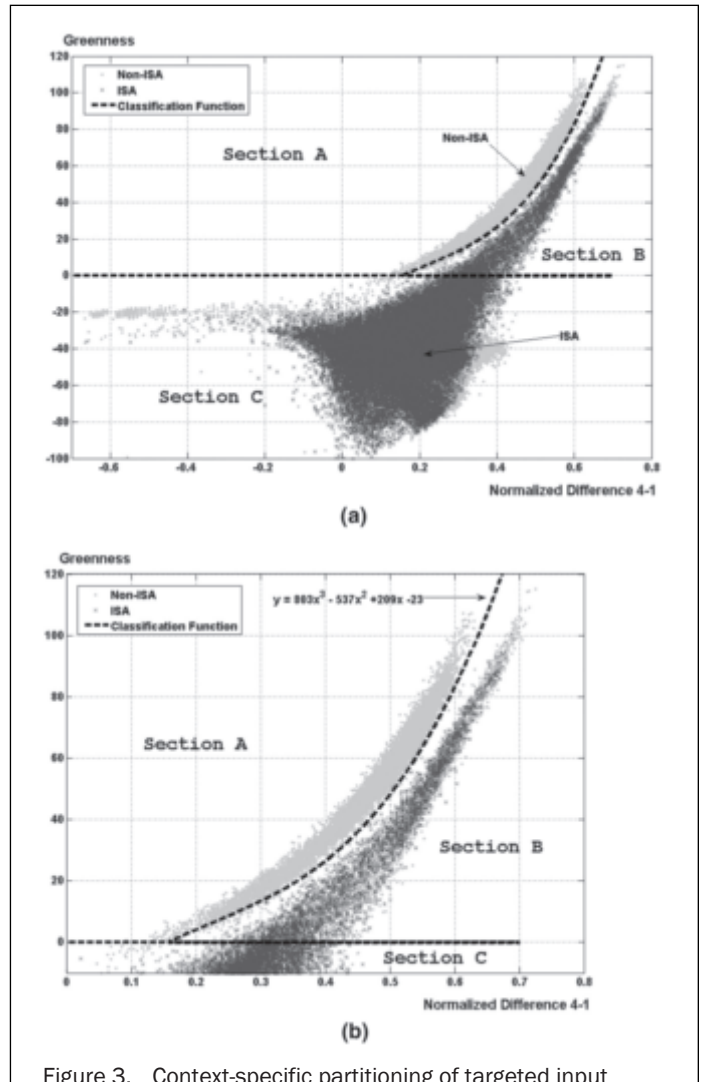

Figure 3. Context-specific partitioning of targeted input space (Node N_1 of Figure 2): (a) Complete input space, and (b) Magnified part of the input space.

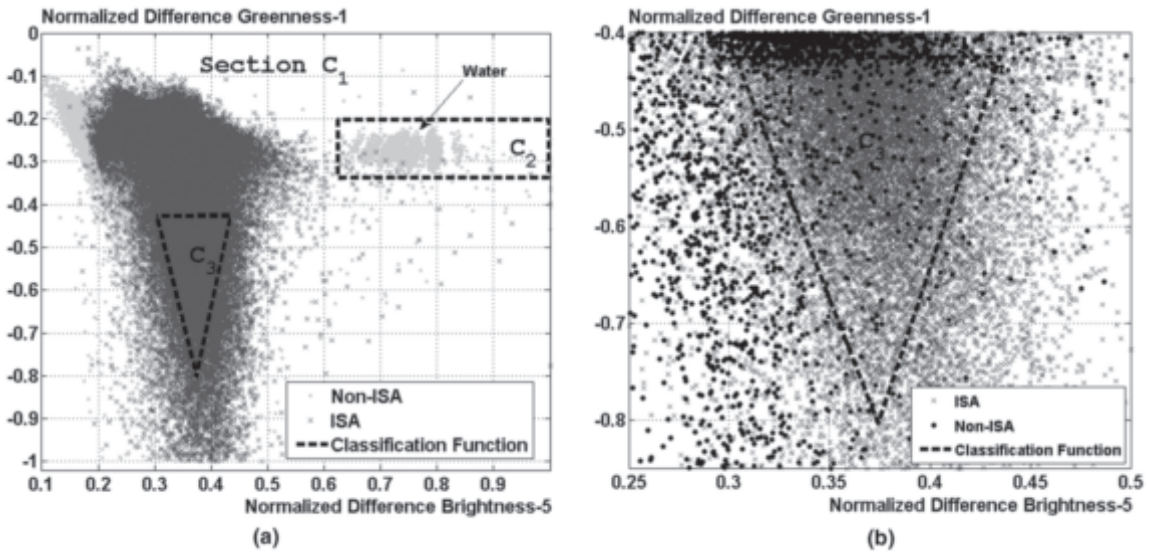

Figure 4. Context-specific partitioning of targeted input space (Node N_2 of Figure 2): (a) Complete input space, and (b) Magnified part of the input space.

accurate in classifying ISA and non-ISA pixels, and then select the most accurate classification from the remaining pairs. We should also mention that not all possible 2D pairs were tested; instead, a subset was evaluated based on human input and computer-generated statistics (e.g., percent of overall ISA pixels that would fall within the 90 percent ratio cells in the 50×50 grid). At the end of the process, the only variables that had been used in 2D separations are those shown in Table 1 plus the additional (and effectively undecipherable) parameters that were generated by processing Table 1 variables in different subsequent algorithms.

Let us examine now the specifics of the identified expert system of Figure 2. Initially, we identify the majority of vegetation and separate that from ISA (Node N_1). Within this task-specific *context* we visualize our training set using the normalized difference of bands 4 and 1, and the Greenness value from the Tasseled cap transformation (Figure 3). We separate the input space into three sections using a line and a polynomial function. Section A contains vegetation directly extracted without further processing. Section B contains a mixture of ISA and vegetation that we need to examine further. This can be done by any appropriate algorithm (e.g., decision trees, neural networks, Bayesian estimation) using any available inputs that fit the specific context of the task. Section B is classified using a support vector machine algorithm with four inputs and a mediocre accuracy of 75 percent. If necessary, section B could be further divided using a different input space to constrain the domains of the algorithms (e.g., grass versus deciduous trees) and potentially increase the classification accuracy. In the Las Vegas sparsely vegetated area and with our ISA identification objectives, such refinements were not deemed applicable.

In addition to sections A and B, we have samples in section C. At this tree node, it is difficult to assign a specific context behind section C (what it contains), so we carry these samples to the next level of our analysis (they are the input samples for node N_2). At the next level (Figure 2, node N_2), we use different inputs but the same methodology as presented above. The analyzed samples are the ones belonging to section C of node N_1. Our classification inputs this time are the normalized difference between Brightness and band 5 for the *X*-axis, and normalized difference between Greenness and band 1 for the *Y*-axis (Figure 4). All the values for Greenness that are contained at this node are negative numbers due to the selection of section C in Figure 3. So, the normalized difference is easily facilitated by using the absolute value of Greenness. We separate this input space into three sections: C_1 contains the context-unidentifiable portion, C_2 contains water samples, and C_3 has ISA samples.

Proceeding to analyze section C_1, we realized that further classification would need more than two inputs due to increased complexity; i.e., our search for parameter pairs with good separation of this population was futile. Visualization and human supervision are substantially more difficult in spaces of more than two dimensions. We dealt with this difficulty by applying an unsupervised algorithm, and then evaluating the results from this algorithm in a low-dimensional space. We used a neural network with six inputs, and projected the accuracy of the network along the brightness input (Figure 5). The algorithm generated accurate results for medium brightness and degraded results under low- or high-brightness conditions. The brightness input space was thus naturally divided into three sections, namely C_{1a}, C_{1b}, and C_{1c}. For C_{1b} we used the tested neural network as the classifier. For the other two sections, we proceeded with further testing.

At this stage, we can assign specific context as to what we are looking for in each of these sections. Researchers in the past were able to identify three types of ISA depending on their brightness (light, medium, dark) but failed to express that explicitly in their classification method (e.g., Hung, 2002). Furthermore, non-ISA features in low-brightness areas are mostly shadow and water, some vegetation, and rare cases of soil. As brightness increases the vegetation presence decreases and soil presence is more evident. The reader should also keep in mind that the majority of vegetation is already absorbed in node N_1.

For section C_{1A} (dark samples), we use a decision tree to separate dark ISA from dark non-ISA. It is a simple classifier with two inputs. This supports our hypothesis that some

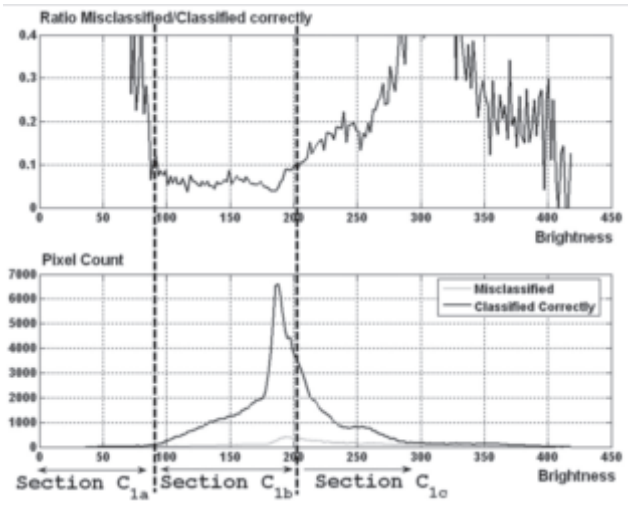

Figure 5. Context-specific partitioning of targeted input space (Node N 3 of Figure 2).

samples are easier to classify than others if a successful hierarchical approach is established. The identified decision tree has adjusted to the lack of complexity of this sub-problem (dark ISA identification).

Figure 5 supports the well-documented spectral resemblance of ISA and specific types of soil represented within the [200, 250] brightness range of section C_{1c}. Since the algorithmic method and inputs used for this section could be completely different than other sections, we can increase the expressiveness of our model by adding complex inputs (e.g., texture statistics and other unusual multi-band inputs). Note that these are computationally expensive inputs that we use only if necessary and for selected parts of the input space.

Samples in section C_{1c} are forwarded to node N_4 (see Figure 2). A simple partitioning takes place at that node using a one-dimensional input, the Normalized Difference between bands 5 and 1 (ND51). Samples with ND51 >0.365 (section C_{1cHigh}) are directly identified as non-ISA pixels. The vast majority of these samples are bright soil. Samples with ND51 ≤ 0.365 are assigned to section C_{1cLow} and require further processing. This is probably the most challenging classification due to the spectral resemblance of bright ISA and soil. We implemented a decision tree to identify ISA within that range. We used six inputs, four of which were based on neighborhood statistics. We stopped the hierarchical analysis at this level since we have reached a predetermined target for accuracy (~90 percent in the whole study area).

Below, we offer two tables to help readers follow this rather complicated but rewarding classification approach. Table 2 summarizes each node, their respective inputs, and the context behind their partitioning. Table 3 identifies key properties for each of the four leaf classifiers.

Results

Statistical Evaluation

We evaluated our method using a dataset created from high-resolution aerial photography. The digitization process assigned a pixel as ISA if any portion of the 30 m cell covered ISA on the high-resolution aerial imagery. The complete evaluation dataset contained 242,469 cells of 30 m resolution; 20,000 pixels were used for training, and the remaining 222,469 pixels were never seen by any algorithms in their training phases and were used for validation. In our tests, we first assessed the incorporation of additional statistics in the classification process and then evaluated our collaborative expert system.

Incorporation of Additional Statistics in the Classification Process

We initially evaluated the expressiveness of the raw Landsat information as opposed to calculated statistics from these six bands (Table 4). We have performed 5,000 simulations for each input combination. Decision trees were trained using a ten-fold cross-validation method and backpropagation neural networks with variable hidden nodes were implemented, both using MATLAB's built-in functions.

For each input combination, we report the ones with the best overall accuracy. Table 4 shows the different input combinations; we started with the six Landsat bands

TABLE 2. NODE CONTEXT TABLE

Node	Input Pixels	Input Part	Context	Dimensions Used	ISA	Non ISA	Output Propagation
1	All Pixels	A	Vegetation	ND41, Greenness		☑	END
1	All Pixels	B	Vegetation ISA	ND41, Greenness			Classifier B
1	*All Pixels*	*C*	*None*	*ND41, Greenness*			*Feeds Node 2*
2	*C Pixels*	*C_1*	*None*	*NDB5, NDaG1*			*Feeds Node 3*
2	C Pixels	C_2	Water	NDB5, NDaG1		☑	END
2	C Pixels	C_3	ISA	NDB5, NDaG1	☑		END
3	C_1 Pixels	C_{1a}	Dark ISA Shadows, Dark Soil/Vegetation	Brightness			Classifier C_{1a}
3	C_1 Pixels	C_{1b}	Medium ISA Soil, Vegetation	Brightness			Classifier C_{1B}
3	*C_1 Pixels*	*C_{1c}*	*Bright ISA Bright Soil*	*Brightness*			*Feeds Node 4*
4	C_{1c} Pixels	C_{1cLow}	Bright ISA Bright Soil	ND51			Classifier C_{1cLow}
4	C_{1c} Pixels	C_{1cHigh}	Bright Soil	ND51		☑	END

TABLE 3. LEAF CLASSIFIER TABLE

Input Samples	Dimensions	Context	Method
B	UBaG1, UBaG5, ND43sd5, ND52sd3	ISA Vegetation	Support Vector Machine
C_{1A}	UNtanB4, ND51	Dark ISA NonISA	Decision Tree
C_{1B}	ND54, ND42, Bsd3, UNtanBWW3, UND1aG15, ND43sd7	Medium ISA Soil, Vegetation	Backpropagation Neural Network
C_{1C_LOW}	NDaG5, UNtanBGWB, ND43sd9, ND45sd3, NDB1sd5, D51sd5	Bright ISA Bright Soil	Decision Tree

and then added the three Tasseled cap transformation statistics (Brightness, Greenness, Wetness) and the Normalized Difference Vegetation Index (NDVI). We followed with the principal components of these two input combinations. At the next step, we randomly selected variables from our 277 input collection (as described in a previous section). We gradually increased the dimensionality from 10 to 15 and then to 20 inputs, and we also assessed the incorporation of neighborhood statistics. Here are our conclusions:

- The incorporation of Tasseled cap transformation statistics and NDVI made no significant difference for either the decision trees or the neural networks.
- The transformation to the principal component (PC) space did not provide any benefits. This is to be expected as PC analysis is not a parameterized approach, i.e., does not consider class outputs in the transformation process.
- The inclusion of additional single pixel statistics had a significant impact on the decision trees as it increased overall accuracy by 1.5 percent to 2 percent. No change was observed for the neural networks.
- The support of neighborhood statistics increased the overall accuracy by 1.6 percent for the decision trees and 0.6 percent to 1 percent for the neural networks. Considering the already achieved accuracy (>90 percent), these increases are substantial and show the value of spatial masks even at moderate resolutions (30 m).

From the above, we can conclude that neural networks clearly outperform decision tress in our application. Furthermore, neural networks' increased mathematical flexibility allows accurate classification without any post-processing of the raw bands, while decision trees clearly benefit from post-processing. Independently of the classification method, neighborhood statistics provided substantial boost.

Comparison of the Collaborative Expert System

We used the same training and validation dataset, and we compared our methodology to decision trees and neural networks (Table 5). We performed two different evaluations: first, we used the exact same inputs identified by our expert system, and then, we selected random combinations of the same number of inputs (23) from the pool of 277 candidates (see fixed and random in Table 5). Each decision tree and

TABLE 4. STATISTICAL COMPARISON FOR INPUT VARIABILITY ON DECISION TREES AND NEURAL NETWORKS

Decision Trees

Dim	Type	Spatial Masks	Ref0-Alg0	Ref0-Alg1	Ref1-Alg0	Ref1-Alg1	User0 (%)	User1 (%)	Prod0 (%)	Prod1 (%)	Overall (%)	Kappa
6	OB	No	146776	10287	17409	47997	89.4	82.4	93.5	73.4	**87.6**	0.69
10	OB_3T_NDVI	No	142673	14390	12364	53042	92.0	78.7	90.8	81.1	**88.0**	0.71
6	PCA(OB)	No	145281	11782	16561	48845	89.8	80.6	92.5	74.7	**87.3**	0.69
6	PCA(OB_3T_NDVI)	No	144825	12238	14953	50453	90.6	80.5	92.2	77.1	**87.8**	0.70
10	Random	No	148788	8275	14751	50655	91.0	86.0	94.7	77.4	**89.6**	0.74
15	Random	No	148128	8935	14202	51204	91.3	85.1	94.3	78.3	**89.6**	0.74
20	Random	No	148471	8592	14444	50962	91.1	85.6	94.5	77.9	**89.6**	0.74
10	Random	Yes	150879	6184	13330	52076	91.9	89.4	96.1	79.6	**91.2**	0.78
15	Random	Yes	149225	7838	11751	53655	92.7	87.3	95.0	82.0	**91.2**	0.78
20	Random	Yes	151335	5728	13829	51577	91.6	90.0	96.4	78.9	**91.2**	0.78

Neural Networks

Dim	Type	Spatial Masks	Ref0-Alg0	Ref0-Alg1	Ref1-Alg0	Ref1-Alg1	User0 (%)	User1 (%)	Prod0 (%)	Prod1 (%)	Overall (%)	Kappa
6	OB	No	148953	8110	12263	53143	92.4	86.8	94.8	81.3	**90.8**	0.78
10	OB_3T_NDVI	No	149262	7801	12547	52859	92.3	87.1	95.0	80.8	**90.9**	0.77
6	PCA(OB)	No	148466	8597	11607	53799	92.8	86.2	94.5	82.3	**90.9**	0.78
6	PCA(OB_3T_NDVI)	No	148988	8075	11871	53535	92.6	86.9	94.9	81.9	**91.0**	0.78
10	Random	No	148518	8545	11248	54158	93.0	86.4	94.6	82.8	**91.1**	0.78
15	Random	No	148841	8222	11760	53646	92.7	86.7	94.8	82.0	**91.0**	0.78
20	Random	No	149355	7708	12631	52775	92.2	87.3	95.1	80.7	**90.9**	0.77
10	Random	Yes	149046	8017	10522	54884	93.4	87.3	94.9	83.9	**91.7**	0.80
15	Random	Yes	148891	8172	10537	54869	93.4	87.0	94.8	83.9	**91.6**	0.80
20	Random	Yes	149127	7936	9963	55443	93.7	87.5	95.0	84.8	**92.0**	0.80

OB: Original six bands, *OB_3T_NDVI*: OB plus three Tasseled cap statistics and NDVI, *Ref*: Reference, *Alg*: Algorithm, *0*: NonISA class, *1*: ISA class, *User*: User's accuracy, *Prod*: Producer's accuracy.

TABLE 5. STATISTICAL COMPARISON OF OUR METHOD WITH DECISION TREES AND NEURAL NETWORKS

Dim	Type	Algorithm	Ref0-Alg0	Ref0-Alg1	Ref1-Alg0	Ref1-Alg1	User0 (%)	User1 (%)	Prod0 (%)	Prod1 (%)	Overall (%)	Kappa
23	Fixed	Decision Tree	148756	8307	14249	51157	91.26	86.03	94.71	78.21	89.86	0.75
23	Random	Decision Tree	150710	6349	13134	52272	91.98	89.17	95.96	79.92	91.24	0.78
23	Fixed	Neural Network	150097	6966	10324	55082	93.56	88.77	95.56	84.22	92.23	0.81
23	Random	Neural Network	148429	8634	9457	55949	94.01	86.63	94.50	85.54	91.87	0.80
23	Fixed	Collaborative Expert System	149138	7925	8899	56507	94.37	87.70	94.95	86.39	**92.44**	**0.82**

neural network was selected as the provider of the best overall accuracy from 20,000 simulations.

Our approach compared favorably to the established methods. The range of overall accuracy increase was from 0.2 percent to 2.6 percent. This is an enhancement considering the limited room for improvement beyond 90 percent accuracy due to other errors (image acquisition and processing). We should emphasize though that the major goal of this paper is to provide the proof-of-concept for collaborative algorithms by showing their applicability with equal or better results.

Internal Accuracy Distribution of the Collaborative Expert System

When complex algorithms such as neural networks are implemented for classification purposes, the accuracy assessment can only be performed at the network output nodes. Variable classification metrics can be provided based on the strength of each output node; however, if a specific output node is found to underperform, the internal network architecture cannot be changed without affecting other output nodes causing retraining for the whole network. This is one of the reasons behind their so-called "black box" behavior and their limited error correcting capabilities.

Our approach on the other hand is harvesting the modeling power of complex mathematical models such as neural networks in a controlled fashion that facilitates advanced yet simple error correcting capabilities. This proposed expert system is composed of eight different classifiers with their corresponding accuracies found in Table 6. We see an interesting imbalance between each algorithm's accuracy; some are higher than 95 percent, while others are below 83 percent (Algorithms 2 and 7). This high variability is desirable because it provides us with the opportunity to further improve accuracy by targeting our efforts in one or two of these underperforming algorithms. Each algorithm's training is independent of the other, so we will not affect high accuracy achieved by the other algorithms, and we do not have to retrain our whole system.

Study Area Generalization

In our hierarchical approach, each of the classifiers is associated with specific accuracy metrics. These metrics are calculated in the calibration process and were provided in Table 6. Upon simulation to the whole study area (partial Landsat scene), our methodology exhibits two advantages resulting from the identification of pixels classified by each of the eight algorithms:

1. We propagate accuracy to specific scene portions, therefore informing non-experts of algorithmic limitations.
2. We relate calibration and simulation datasets to identify disproportional variations among them. This acts as a guide if further evaluation pixels are acquired.

Table 7 shows the simulation of our approach to the Las Vegas study scene. Clearly, segment A was overemphasized in our calibration dataset (justifiably since we used a vegetation dataset outside Las Vegas for training, but no vegetation exists in Las Vegas), while segment C_{1cLow} could use additional training pixels to represent better the study area. In terms of overall accuracy the 92.44 percent observed in calibration would propagate into a 90.99 percent in the entire scene simulation. Typical remote sensing classifiers do not distinguish between calibration and simulation accuracy.

The reader may notice the small scene contribution of the first five classifiers. It is true that they do not contribute considerably to the overall solution (~8 percent),

TABLE 6. ACCURACY DISTRIBUTION WITHIN OUR COLLABORATIVE EXPERT SYSTEM

Algorithm	Segment	Ref0-Alg0	Ref0-Alg1	Ref1-Alg0	Ref1-Alg1	User0 (%)	User1 (%)	Prod0 (%)	Prod1 (%)	Overall (%)	Kappa
1	A	9773	N/A	1	N/A	99.99	N/A	N/A	N/A	99.99	N/A
2	B	1788	592	663	1883	72.95	76.08	75.13	73.96	74.52	0.49
3	C_2	3832	N/A	44	N/A	98.86	N/A	N/A	N/A	98.86	N/A
4	C_3	N/A	182	N/A	10123	N/A	98.23	N/A	N/A	98.23	N/A
5	C_{1a}	1039	53	55	263	94.97	83.23	95.15	82.70	92.34	0.78
6	C_{1b}	90382	3112	2752	31728	97.05	91.07	96.67	92.02	95.42	0.88
7	C_{1cLow}	28047	3986	4780	12510	85.44	75.84	87.56	72.35	82.23	0.61
8	C_{1cHigh}	14277	N/A	604	N/A	95.94	N/A	N/A	N/A	95.94	N/A
Total	All	149138	7925	8899	56507	94.37	87.70	94.95	86.39	92.44	0.82

N/A: Not applicable and it relates to the fact that those algorithms extracted directly a single class.

TABLE 7. PROJECTING ACCURACY DISTRIBUTION FROM CALIBRATION TO SIMULATION

		Calibration		*Simulation*
Algorithm	Segment	Segment Accuracy (%)	Segment Portion (%)	Scene Portion (%)
1	A	99.99	4.39	0.01
2	B	74.52	2.21	1.62
3	C_2	98.86	1.74	1.13
4	C_3	98.23	4.63	3.47
5	C_{1a}	92.34	0.63	0.51
6	C_{1b}	95.42	57.52	49.46
7	C_{1cLow}	82.23	22.17	32.42
8	C_{1cHigh}	95.94	6.69	11.39

since vegetation and water presence in our Las Vegas study area is limited. However, the purpose of this paper is to establish a unified framework for context-specific multi-method classifiers. In other settings these classifiers may process a higher percentage of pixels. For example, we are currently implementing our approach in Syracuse, New York where vegetation and water have a stronger presence.

A representation of the impervious surface classification results from our Las Vegas pilot study area follows (Figure 6). In addition to the accurate classification, an accuracy layer is produced (Figure 7). This accuracy layer is a major step towards support of non-experts, algorithmic refinement and targeted acquisition of external datasets. This layer was produced by identifying which classifier (A, B, C_{1a}, C_{1b}, C_{1cLow}, C_{1cHigh}, C_2, or C_3) was used to categorize each pixel in the hierarchical structure of Figure 2. Each classifier has an associated accuracy metric (see Table 6) and that is the accuracy we present to the user.

The accuracy for the whole study area is projected to be 90.99 percent. Depending on the application focus more specific accuracy metrics can be produced. For example, if

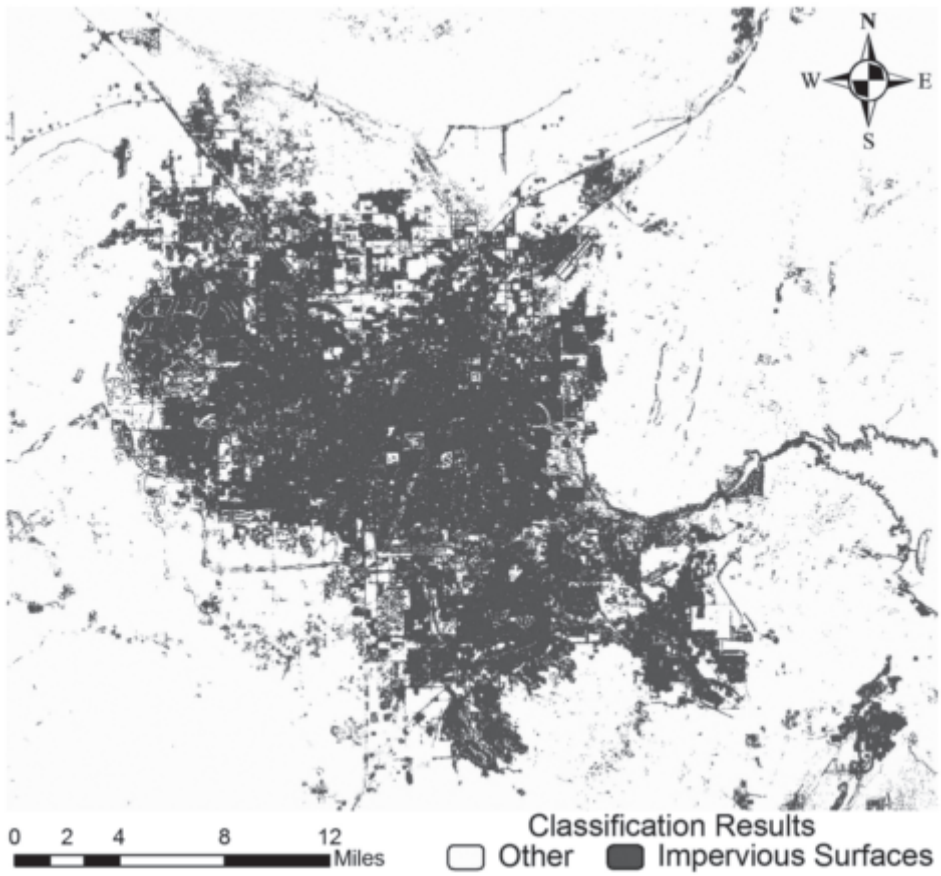

Figure 6. ISA Classification results.

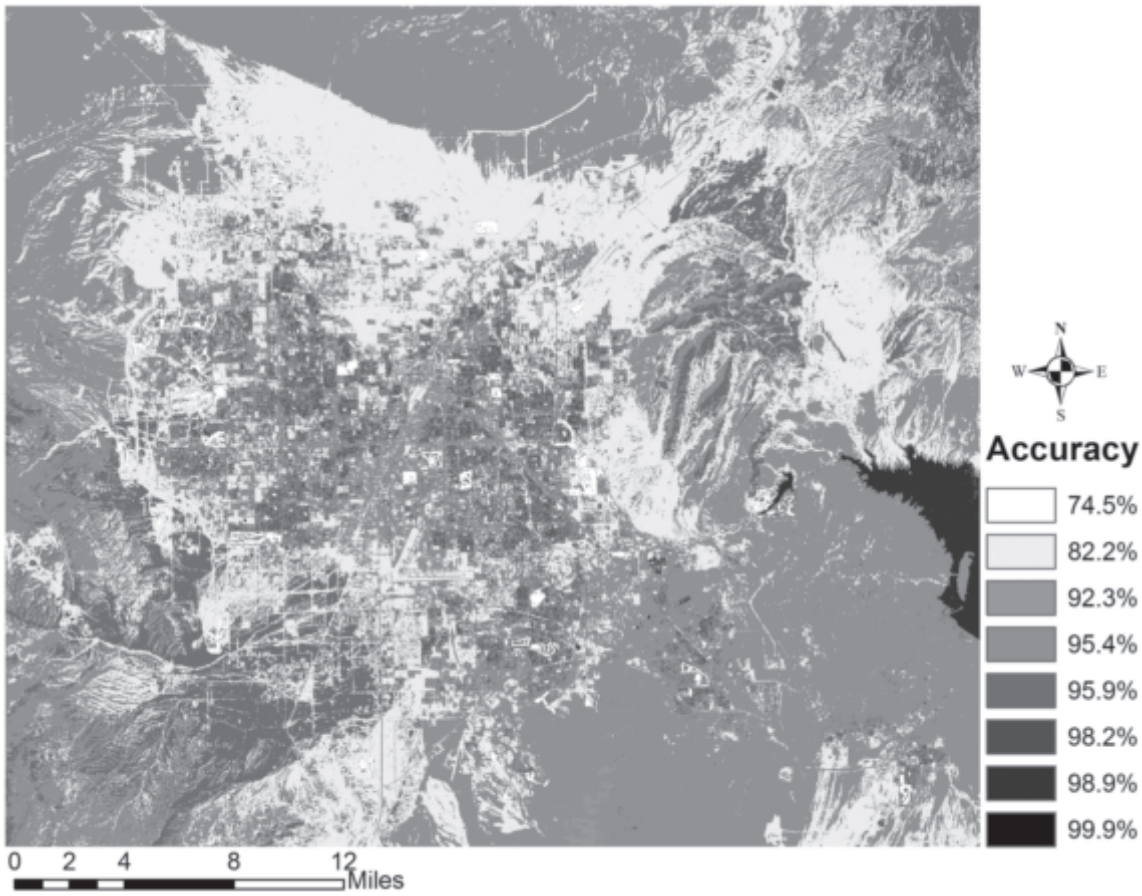

Figure 7. Classification accuracy of each algorithm. A color version of this figure is available at the ASPRS website: **www.asprs.org** .

we split the area in four equal quarters the accuracy would be 90.23 percent, 89.77 percent, 92.93 percent and 91.01 percent, respectively, moving clockwise from the northwestern quarter.

Discussion and Conclusions

The methodology presented in this paper is a complex expert-based system, and in its current form, it targets multi-scene, large-scale applications rather than single scene classification. Our goal is not to present yet another single-thread classifier; instead, we strive to establish a framework for collaborative algorithms. This in turn would allow incremental algorithmic improvements instead of isolated methods addressing problems already partially solved.

Our work is not the first to produce pixel-based accuracy estimates of the classification (Carpenter *et al.*, 1999, Pontius, 2000, Alimohammadi *et al.*, 2004, Liu *et al.*, 2004, Aires *et al.*, 2004). We build on these limited efforts by adding an important component behind the concept of having multiple classifiers merged in a unified solution, the concept of targeted, context-specific classifiers. By decomposing the assigned task (e.g., ISA versus non-ISA classification) into smaller, more manageable sub-tasks (e.g., dark-ISA versus dark-soil) several benefits of our hierarchical context-specific approach can be identified. Starting with internal algorithmic behavior we list the following.

Support for Targeted Classification Inputs

Combination of various inputs on-demand allows a more expressive methodology than a typical single-method classification using fixed inputs. Inputs are selected to distinguish classes in limited parts of the input space, therefore facilitating increased discriminatory power. This by itself does not guarantee higher accuracy, but it increases the potential for higher accuracy, if the sub-problem identification is done properly. We have also demonstrated how part of the problem can be solved in a low dimensional input space (one- or two-dimensional inputs) where visualizations are easy and expressive in terms of assigning context.

Support for Different Algorithmic Methodologies

Another characteristic of our approach is that it is not restricted to a single classification methodology (e.g., decision trees, neural networks). In this demonstration, we have used polynomials, rectangles, and triangles in low 2D spaces combined with backpropagation neural networks, decision trees, and support vector machines in spaces of higher dimensionality. Arbitrary mixing of different learning approaches is supported, if desired. Therefore, the complexity of each process within our method adjusts to the complexity of the underlying sub-task. As we demonstrated in Figure 3, a simple polynomial can distinguish some types of vegetation, while a neural network is necessary to differentiate ISA and some soil types. We should

also note that different methods could work competitively in addition to collaboratively. Different algorithmic methods can be implemented for the same sub-task to identify the prevailing method.

Moving beyond benefits to algorithmic performance, our methodology extends current state of the art through detailed accuracy analysis. An important characteristic is that each process is tied to specific accuracy; therefore a different accuracy measure is assigned from each classifier giving a more representative, spatially explicit accuracy value inherited from the process used. This results in numerous advantages when simulating on a new spatial scene:

Identification of Areas for Algorithmic Improvement and/or Ancillary Data Acquisition

We know in advance the portion of pixels falling within each module (process). This translates into knowledge of problematic areas before even the actual simulation takes place. For example, after filtering takes place and mostly soil and ISA remain, we know how many pixels fall within the problematic [200, 250] brightness range. Using this as a guide, we can decide whether we need to improve algorithmic performance through additional training and model testing. Furthermore, if spectral discrimination is not feasible a decision can be made to acquire ancillary data (e.g., high-resolution imagery, lidar, or census data) that may enhance our final product. Our approach can act as the basis for image/data fusion algorithms.

Evaluation of Training Site Representation in Overall Scene

Traditional classifiers do not establish a within class relationship between training and simulation datasets. Our methodology not only propagates classification accuracy as mentioned above but also establishes linkages for every class expressing variability in the training and simulation datasets. For example, Table 7 shows multiple corresponding training and simulation portions within each class. This can be used as a guide for future training data acquisition.

Support for Multiple Specialists Engagement

Our approach can establish a framework where the general problem can be decomposed to multiple sub-problems and subsequently assign specific sub-problems to different scientists and/or research groups. More specifically, the structured examination of error within the parameter space at each level of the decision tree provides a basis for engaging appropriate specialists in identifying methods for the next stage of classification (e.g., vegetation versus soil versus ISA classifiers). By the same token, multiple scientists can compete for a sub-problem classifier using different machine learning methods.

Straightforward Assessment of Results by Non-experts

Remote sensing products often act as a critical input to studies from a variety of disciplines. However, non-experts have high expectations from remote sensing products without realizing potential sensor and modeling limitations. Therefore, there is a clear need to incorporate advanced accuracy metrics associated with remote sensing products that express usefulness and limitations of incorporated methodologies. As our results section has demonstrated, accuracy maps are a natural product of our multi-method framework.

Simulation Speed

We reserve the computationally expensive inputs (e.g., texture-based statistics) only for sub-tasks where conventional statistics (e.g., normalized band differences) do not provide adequate separation. By doing so, the computationally expensive inputs are calculated only if necessary and for selected parts of the imagery, therefore accelerating the simulation speed. This is not a major concern with current computational power when simulating a single satellite scene, but we envision utilization of this approach into larger datasets at multiple time instances. Furthermore, several classifiers can be employed in a parallel fashion to accelerate simulation, utilizing multi-core processors.

In addition to the aforementioned benefits, our methodology currently exhibits a significant drawback. As an expert-based system, the selection of nodes and the assignment of context behind their partitioning are not currently performed in an automated manner. We have developed some statistical tools that help us with exploratory analysis in low-dimensional space, but still substantial human involvement is necessary. This paper is acting as the proof-of-concept and establishes the baseline for needed improvements. We are currently testing incorporation of a brute force approach to automatically select the best combination of algorithms. In the future, it would also be interesting to see if seasonal variations can be constrained and rectified in selected algorithms.

Acknowledgments

At the early stage, this research was supported by the U.S. Geological Survey through a National Academies of Science Postdoctoral Award for Dr. Mountrakis. Further support was provided by NSF Award No. 0648393, NASA's New Investigator Program, and the Syracuse Center of Excellence CARTI Program. We would like to thank Colin Homer and George Xian from the USGS for sharing their training datasets.

References

Aires, F., C. Prigent, and W.B. Rossow, 2004. Neural network uncertainty assessment using Bayesian statistics: A remote sensing application, *Neural Computation*, 16:2415–2458.

Aitkenhead M.J., and R. Dyer, 2007. Improving land-cover classification using recognition threshold neural networks, *Photogrammetric Engineering & Remote Sensing*, 73(4):413–421.

Alimohammadi, A., H.R., Rabiei, and P.Z. Firouzabadi, 2004. A new approach for modeling uncertainty in remote sensing change detection process, *Proceedings of the 12th International Conference on Geoinformatics - Geospatial Information Research: Bridging the Pacific and Atlantic*, 07–09 June, University of Gavle, Sweden, pp. 503–508.

Bauer, M.E., and L. Brian, 2005. Estimation, mapping and change analysis of impervious surface area by landsat remote sensing, *Proceedings of Pecora 16 - Global Priorities in Land Remote Sensing*, 23–27 October, Sioux Falls, South Dakota.

Berry, B.J.L., and F.E. Horton, 1974. *Urban Environmental Management: Planning for Pollution Control*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Breiman, L., 1996. Bagging predictors, *Machine Learning*, 24(2):123–140.

Canters, F., 1997. Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification, *Photogrammetric Engineering & Remote Sensing*, 63(4):403–414.

Carpenter, G.A., S. Gopal, S. Macomber, S. Martens, C.E. Woodcock, and J. Franklin, 1999. A neural network method for efficient vegetation mapping, *Remote Sensing of Environment*, 70:326–338.

Coe, S.E., M. Alberti, J.A. Hepinstall, and R. Coburn, 2005. A hybrid approach to detecting impervious surface at multiple-scale, *Proceedings of the ISPRS WG VII/1 - Human Settlements and Impact Analysis - 3rd International Symposium Remote Sensing and Data Fusion Over Urban Areas (URBAN 2005)*, 14–16 March, Tempe, Arizona.

Flanagan, M., and D.L. Civco, 2001. Subpixel impervious surface mapping, *Proceedings of 2001 ASPRS Annual Convention*, 23–27 April, St. Louis, Missouri, unpaginated CD-ROM.

Foody, G.M., N.A. Campbell, N.M. Trodd, and T.F. Wood, 1992. Derivation and applications of probabilistic measures of class membership from maximum-likelihood classification, *Photogrammetric Engineering & Remote Sensing*, 58(9):1335–1341.

Forman, R.T.T., 2003. *Road Ecology: Science and Solutions*, Island Press, Washington, D.C.

Forman, R.T.T., and L.E. Alexander, 1998. Roads and their major ecological effects, *Annual Review of Ecological Systems*, 29:207–231.

Forman, R.T.T., and R.D. Deblinger, 2000. The ecological road-effect zone of a Massachusetts (U.S.A.) suburban highway, *Conservation Biology*, 14(1):36–46.

Forster, B.C., 1980. Urban residential ground cover using Landsat digital data, *Photogrammetric Engineering*, 46(4):547–558.

Hansen, L.K., and P. Salamon, 1990. Neural network ensembles, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:993–1001.

Hodgson, M.E., J.R. Jensen, J. Tullis, K. Riordan, and R. Archer, 2003. Synergistic use of lidar and color aerial photography for mapping urban parcel imperviousness, *Photogrammetric Engineering & Remote Sensing*, 69(9):973–980.

Homer, C., M. Coan, C. Huang, L. Yang, and B. Wylie, 2004. Development of a 2001 National Land-Cover Database for the United States, *Photogrammetric Engineering & Remote Sensing*, 70(7):829–840.

Hung, M., 2002. Urban land cover analysis from satellite images, *Proceedings of Pecora 15-Land Satellite Information IV/ISPRS Commission I/FIEOS 2002 Conference*, 10–15 November, Denver, Colorado.

Jain, A.K., R.P.W. Duin, and J.C. Mao, 2000. Statistical pattern recognition, A review, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37.

Jensen, J.R., 2005. *Introductory Digital Image Processing: A Remote Sensing Perspective*, Third edition, Prentice Hall Series in Geographic Information Science, Upper Saddle River, New Jersey.

Ji, M.H., and J.R. Jensen, 1999. Effectiveness of subpixel analysis in detecting and quantifying urban imperviousness from Landsat Thematic Mapper imagery, *Geocarto International*, 14(4):31–39.

Kennen, J.G., 1999. Relation Of macroinvertebrate community impairment to catchment characteristics in New Jersey streams, *Journal of the American Water Resources Association*, 35(4):939–955.

Krogh, A., and J. Vedelsby, 1995. Neural network ensembles cross validation, and active learning, *Advances in Neural Information Processing Systems* (G. Tesauro, D. Touretzky, and T. Leen, editors), 7, MIT Press, Cambridge, Massachusetts, pp. 238.

Liu,W.G., S. Gopal, and C.E. Woodcock, 2004. Uncertainty and confidence in land-cover classification using a hybrid classifier approach, *Photogrammetric Engineering & Remote Sensing*, 70(8):963–971.

Mastuyama T., 1989. Expert systems for image processing: Knowledge-based composition of image analysis processes, *Computer Vision, Graphics, and Image Processing*, 48:22–49.

McCauley, S., and S.J. Goetz, 2004. Mapping residential density patterns using Landsat imagery and a decision-tree classifier, *International Journal of Remote Sensing*, 25 (6):1077–1094.

Mountrakis, G., 2008. Next generation classifiers: Focusing on integration frameworks, *Photogrammetric Engineering & Remote Sensing*, 74(10):1178–1180.

Perrone, M., 1992. A soft-competitive splitting rule for adaptive treestructured neural networks, *Proceedings of the International Joint Conference on Neural Networks*, Baltimore, Maryland, pp. 689–693.

Pontius, R.G., 2000. Quantification error versus location error in comparison of categorical maps, *Photogrammetric Engineering & Remote Sensing*, 66(8):1011–1016.

Powell, R.L., D.A. Roberts, P.E. Dennison, and L.L. Hess, 2007. Sub-pixel mapping of urban land cover using multiple end-member spectral mixture analysis: Manaus, Brazil, *Remote Sensing of Environment*, 106(2):253–267.

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(12):2165–2185.

Small, C., 2002. Multitemporal analysis of urban reflectance, *Remote Sensing of Environment*, 81:427–442.

Spellerberg, I.F., 2002. *Ecological Effects of Roads, Land Reconstruction and Management, Vol. 2*, Science Publishers, Enfield, New Hampshire.

Steele, B.M., 2000. Combining multiple classifiers: An application using spatial and remotely sensed information for land cover type mapping, *Remote Sensing of Environment*, 74:545–556.

Steele, B.M., J.C. Winne, and R.L. Redmond, 1998. Estimation and mapping of misclassification probabilities for thematic land cover maps, *Remote Sensing of Environment*, 66(2):192–202.

Trombulak, S.C., and C.A. Frissell, 2000. Review of ecological effects of roads on terrestrial and aquatic communities, *Conservation Biology: The Journal of the Society for Conservation Biology*, 14 (1):18–30.

Tso, B., and P.M. Mather, 2001. *Classification Methods for Remotely Sensed Data*, Taylor and Francis, London, U.K.

Wang, Y.Q., X. Zhang, and W. Lampa, 2000. Improvement of spatial accuracy in natural resources mapping using multisensor remote sensing and multisource spatial data, *Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, July, Amsterdam, Netherlands, pp. 723–730.

Ward, D., S.R. Phinn, and A.T. Murry, 2000. Monitoring growth in rapidly urbanized areas using remotely sensed data, *Professional Geographer*, 52(3):371–386.

Weng, Q., 2007. *Remote Sensing of Impervious Surfaces*, CRC Press, Boca Raton, Florida.

Wolpert, D.H., 1992. Stacked generalization, *Neural Networks*, 5:241–259.

Wu, C., and A.T. Murray, 2003. Estimating impervious surface distribution by spectral mixture analysis, *Remote Sensing of Environment*, 84:493–505.

Yang, L., C. Huang, C.G. Homer, B.K. Wylie, and M.J. Coan, 2003. An approach for mapping large-area impervious surfaces: Synergistic use of Landsat-7 ETM+ and high spatial resolution imagery, *Canadian Journal of Remote Sensing*, 29:230–240.

Yang, X., 2006. Estimating landscape imperviousness index from satellite imagery, *IEEE Geoscience and Remote Sensing Letters*, 3(1):6–9.

Yuan, F., and M.E. Bauer, 2006. Mapping impervious surface area using high resolution imagery: A comparison of object-based and per pixel classification, *Proceedings of the ASPRS 2006 Annual Conference*, Reno, Nevada, unpaginated CD-ROM.