Enhanced Binary Tree Genetic Algorithm for Automatic Land Cover Classification

Kannappan Palaniappan, Feng Zhu, Xinhua Zhuang, Yunxin Zhao and Andrew Blanchard[†]

Multimedia Communication and Visualization Laboratory
Department of Computer Engineering and Computer Science
[†]Department of Electrical Engineering
University of Missouri-Columbia, MO 65211

ABSTRACT: The development of automatic land cover classification maps using validated statewide datasets supported by state and federal agencies are becoming an important tool for monitoring change, planning, and land use impact assessment. Accurate and fast classification algorithms, adaptive to different data sources and scales, will facilitate cost effective routine updating of land cover maps using current satellite imagery to monitor and detect a broad array of land cover phenomena. A new binary decision tree classifier incorporating an evolutionary genetic learning algorithm is proposed for land cover classification. The new classifier referred to as the Enhanced Binary Tree Genetic Algorithm (BTGA+) has been applied to automatically classify tens of millions of pixels in two full scenes of Landsat TM data using multispectral multitemporal radiance features. The BTGA+ classifier can assign pixels into eight land cover categories for Landsat TM scenes in central Missouri with nearly 90% classification accuracy.

INTRODUCTION

The Multi-Resolution Land Characteristics (MRLC) consortium has been instrumental in coordinating the development of a complete land cover for the conterminous United States and a foundational land characteristics data base covering global regions for agency requirements [1]. The MRLC involves four federal agencies and six national environmental monitoring programs including the US Environmental Protection Agency (EMAP, NALC), US Geological Survey (GAP, NALC, NAWQA), National Oceanic and Atmospheric Administration (C-CAP), US Forest Service (RSAC) in collaboration with the EROS Data Center. Most states did not have current maps of land cover and the GAP analysis project was the first state- and national-level effort to produce multiresolution land cover maps using Landsat Thematic Mapper (TM) satellite imagery predominantly for the years 1991 to 1993 [2]. Supervised clustering and classification of Landsat TM data refined with other data sources including visual photointerpretation (satellite and aerial), air video, maps and field observations was used to produce the final land cover products [3]. The GAP project is ongoing with nearly 80% of the states anticipated to finish by 2000 and completion of all states including Alaska and Hawaii in 2005. As part of the Missouri GAP project, an initial statewide land cover classification was completed in 1999 by the Missouri Resource Assessment Partnership (MoRAP) [4].

Due to the high cost of producing land cover maps ranging from several hundred thousand to millions of dollars, the GAP products are intended to be updated only every 5 to 10 years depending on government needs and funding. Automatic algorithms would enable more frequent updating of maps, with adaptive (application or region specific) land cover land use categories, facilitate change detection, assess landscape management practices, and readily incorporate different satellite instruments and ground sampling resolutions. We have developed an automatic land cover classification algorithm using the MoRAP land cover datasets in order to evaluate the accuracy, scalability, cost effectiveness, and computational resource requirements needed to update state and regional land cover maps.

Land cover maps are required by land managers, planners, scientists and policy makers. The principle behind the GAP project was the insight that managing biological diversity is less expensive to society than responding to species extinction or extirpation [2]. In addition to GAP goals such as biodiversity conservation, habitat mapping, species interrelationships, ecological risk assessment, or sustainability of natural resources, land cover maps provide a basic geographic information layer. Land cover maps are useful for land resource management, monitoring coastal land cover, forest cover monitoring, global deforestation tracking, urbanization impact assessment, erosion monitoring, environmental impact assessment water quality factors, and holistic multiscale landscape metrics.

In this paper we describe a decision tree-based supervised classification scheme known as the improved BTGA+ classifier [5] for automatic production of state-wide land cover classifications using labeled training data. The BTGA+ classifier (a) handles large datasets efficiently, (b) uses a fast (integer) GA learning procedure for determining piecewise linear boundaries, (c) supports better initialization for the GA that improves classification accuracy significantly, and (d) has high discrimination and generalization power for compact trees. Automatic algorithms for land cover mapping using supervised classification can be applied or extended to related classification tasks. Such as land use mapping in urban areas utilizing categories correlated with surface permeability for hydrology and storm water management, vegetation species identification, crop identification, crop stress, etc.

Land cover maps focus on classifying vegetation into several distinguishable categories. The vegetation at a

specific site is highly dependent on the interaction of climate, soil type, elevation, slope, latitude, local biota, human impact, etc. The physiognomic characteristics for determining land cover classes include life form, spacing, phenology, and morphology. Land cover classes used in some GAP projects are based on a national vegetation classification system by the Federal Geographic Data Committee and is related to international terrestrial cover classifications [3]. Classes for climax vegetation include six primary categories: forests, woodlands, shrublands, dwarf-shrublands, grasslands, barren.

Automatic land cover classification requires quantifiable differentiation of vegetation classes using primarily spectral features from remotely sensed imagery. Additionally, vegetation classes used by neighboring geographical regions (i.e. states) should be compatible. The interagency land cover group advising MoRAP has identified an hierarchical classification scheme. This study uses only the eight primary classes in this scheme: urban, cropland, forest (tree canopy >80%), woodland (tree canopy 20-80%, may include thin forest, old field with trees, pasture with trees, etc.), shrubland (>25% shrub cover), herbaceous (>75% cover, graminoid-forb), water, and barren/sparsely vegetated.

Decision tree-based classification algorithms are a successful class of machine learning approaches that have been applied to automatic land cover classification at several scales [6][7]. Remote sensing data classification using decision trees have substantial advantages because of their flexibility, intuitive simplicity, computational efficiency, ability to handle missing or noisy data, and do not require explicit class parameter, model or distribution estimation. Some decision trees are created manually, like a classification key, when the number of features is few and the classes are easily are distinguishable. More commonly, the classification structure defined by a decision tree is estimated from training data using a learning procedure [8][9]. A variety of other approaches for land cover classification have also been investigated including artificial neural networks, Bayesian classifiers, maximum likelihood, etc. [10] The remainder of the paper describes the BTGA+ approach and experimental results.

ENHANCED BINARY TREE GENETIC ALGORITHM (BTGA+) CLASSIFIER

Conventional decision tree algorithms such as the parallelepiped classifier use one feature at a time and search for optimal axes parallel separating planes at each node of a decision tree. The BTGA+ approach uses piecewise linear decision boundaries to better approximate complex class boundaries [5]. Associated with each decision node in the binary tree is a hyperplane that divides the multidimensional space into two separate regions. A genetic algorithm is used to search for optimal decision boundaries using labeled training data. Node splitting continues until one of several stopping conditions is satisfied, such as the minimum terminal node size, minimum node impurity or minimum feature variance. The binary tree maintains a hierarchical

sequence of weight vectors in the non-terminal or decision nodes corresponding to each piecewise linear decision boundary. The plurality or majority rule is used to label the terminal or leaf node with the largest population class.

The essence of the binary decision tree classifier is to split the training data at each node into two subsets, so that the data after splitting (i.e. in each subtree) becomes purer, using a suitable measure of class homogeneity, in comparison to the parent node. Let $X_t = \{x_1, x_2, ..., x_{N_t}\}$ be the training data arriving at the current node t with each sample belonging to one of c classes. The ith observation, $x_i = \{x_{1i}, x_{2i}, ..., x_{d_i}\}^T$, in the training data with d features, can be assigned to one of two subsets $X_{tL}(w)$ and $X_{tR}(w)$, for the left and right children nodes respectively, using the linear decision function $x_i^T \tilde{w} + w_0 \leq 0$. Denoting the impurity of node t as $I(X_t)$, the *impurity reduction* gained by splitting node t is calculated by subtracting the weighted average subset impurities,

$$\Delta I(X_t, w) = I(X_t) - p(X_{tL}(w) / X_t)(I(X_{tL}(w)) - p(X_{tR}(w) / X_t)I(X_{tR}(w))$$

where $p(X_{tL} / X_t) = \# X_{tL} / \# X_t$. The objective is to find a weight vector $w = \{w_0, \tilde{w}\} = \{w_0, w_1, ..., w_d\}^T$ that maximizes the impurity reduction for a given node split.

Three different impurity measures were tested [9]. The *Gini index* of impurity,

$$I(X_t) = \sum_{i=1}^{c} \sum_{j \neq i} p(C_i / X_t) p(C_j / X_t)$$

The *information gain*,

$$I(X_t) = -\sum_{i=1}^{c} p(C_i / X_t) \log p(C_i / X_t)$$

The *Twoing rule* uses the expression for impurity reduction:

$$\Delta I(X_t) = \frac{\# X_{tL} \# X_{tR}}{4 * \# X_t^2} \left[\sum_{i=1}^{c} \left| \frac{\# X_{i,L}}{\# X_L} - \frac{\# X_{i,R}}{\# X_R} \right| \right]^2$$

The Gini index-based impurity reduction is the simplest to evaluate provided the best or comparable results.

The optimal decision function w for each node is determined using a genetic algorithm which provides a global optimization procedure. The chromosomes in the GA represent the weight vector. For each generation, chromosomes compete with each other using impurity reduction as the genetic fitness function. The chromosomes with larger impurity reduction have a higher chance of being included in the recombination pool to form the next generation. The GA terminates after a maximum number generations or convergence of the fitness function. The decision tree is grown until one of the stopping conditions is met; either a minimum pre-defined terminal node size criterion (TNSC) or a minimum impurity reduction criterion (IRC). The overall algorithm is summarized in Figure 1.

Bayesian Initialization with Randomization

Chromosome initialization (i.e. initial weight vectors) is crucial to the quality of the decision boundary and the speed

with which the GA converges. Using random initialization, the initial chromosomes may start off quite far away from the global optimum and even after substantial search the decision boundary is often a local optimum. A Bayesian initialization with randomization is used to initialize the chromosome population. This leads to faster convergence and a more globally optimal weight vector that leads to higher impurity reduction and lower misclassification errors. The Bayesian intialization uses the common covariance weighted vector between the means of the dominant class, C_0, and the rest of the node samples, C_1, with $\tilde{w}^T = \Sigma^{-1}(\mu_0 - \mu_1)$ and

$$w_0 = -\frac{1}{2}(\mu_0^T \Sigma^{-1} \mu_0 - \mu_1^T \Sigma^{-1} \mu_1) + \ln \frac{p(C_0)}{p(C_1)}$$

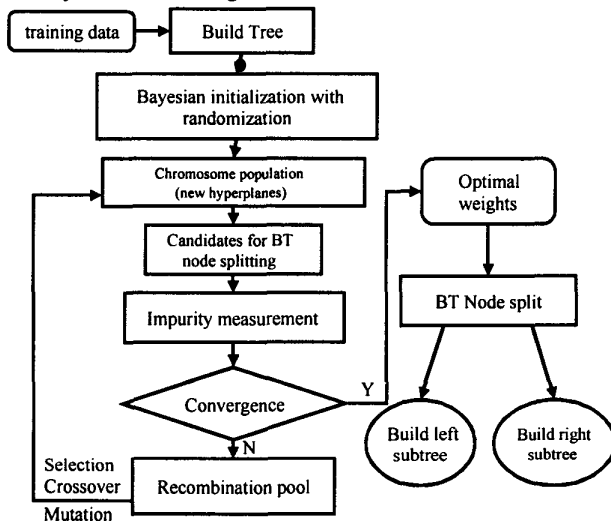Each component of the weight vector is perturbed using Cauchy white noise to generate additional chromosomes.

Figure 1. Steps for depth first BTGA+ classifier construction.

EXPERIMENTAL RESULTS

The state of Missouri is situated in the middle of the conterminous US, is bounded by seven states, covers 69,674 square miles, has an elevation difference of 443 meters, with only 6.7% of public lands, and the previous land cover land use layer at finer than 40 acre mapping units was completed in the early 1970's [4]. The GAP land cover project, among the first in the midwest, used thirty 185km x 185km Landsat TM scenes (fifteen scenes at two dates) between 1991 and 1993; some scenes had only partial coverage of Missouri. The two Landsat scenes selected for testing the BTGA classifier are Path 25 Row 33 and Path 25 Row 34 referred to as Scenes 2533 and 2534 respectively. The Landsat TM sensor has 7 channels covering the visible, near infrared, and thermal infrared bands at 30m, 30m and 120m nominal ground resolution with a 16 day revisit period. The mid-infrared bands make the TM sensor among the best for vegetation mapping. The thermal channel was not used in the supervised or automatic land cover classifications. The unsupervised

spectral clustering combined with supervised class labeling was performed, interpreted and refined by MoRAP. This involved selection of the most informative bands, spatial sectioning by ecological regions, identification of sixty spectral classes using multitemporal TM data, aggregation of spectral groups into eleven land cover categories, post-classification mosaicing, land cover verification using 1:40,000 National Aerial Photography Program (NAPP) photography [4]. The Missouri landscape is 70% agricultural (row crop or pasture) and the forested area primarily rural. The class distributions for the state are: deciduous forest (28.1%), evergreen forest (1.6%), mixed forest (1.8%), deciduous woodland (0.8%), deciduous shrubland (<0.1%), grassland (33.1%), row crop (28.6%), urban (3.7%), open water (2.2%). The urban and water categories were initially based on map-based masks and were not based on spectral clustering (a later Phase II land cover map rectified this by using spectral information) and a barren class was later added. Several ground truth assessments determined that user accuracy for the forest class increased from 73% to 88% by grouping the forest classes into one category [4]. Using this revised eight category classification the MoRAP land cover map is estimated to have an overall accuracy of 78% [4].

Scene 2533 is 7226 x 6524 pixels, scene 2534 is 7707 x 6867 pixels and the subimage AOI1 located at offset (3500, 3500) from the upper left corner of scene 2533 have class distributions as shown in Table 1. Seasonal imagery from two dates (3May92, 24Sep92) and six channels per date for a total of 12 spectral features were used as the features for classification. Pixel-based land cover class labels from the MoRAP map were used to train the BTGA+ classifier under the assumption that the MoRAP land cover assignment is entirely correct. The BTGA+ classifier is evaluated using the three datasets in Table 1 for nearly 70 million pixels.

Class ID	Land cover classes	Scene 2534 Hist.(%)	Scene 2533 Hist.(%)	AOI1 512x512 Hist.(%)
0	Urban	0.73	0.55	0.03
1	Cropland	0.62	30.46	19.57
2	Shrubland	0.01	0.26	0.0
3	Open water	2.01	1.69	4.36
4	Barren/sparsely veg.	0.11	0.15	0.0
5	Forest	40.95	26.12	37.95
6	Woodland	6.93	4.04	5.56
7	Herbaceous	48.63	36.73	32.53

Table 1. The 8 major land cover classes and their distribution in two scenes and within a small area of interest in 2533.

The three impurity reduction measurements were tested and produced similar overall accuracies. So the Gini index was used since it can be efficiently evaluated. The optimal values for several of the BTGA+ parameters were determined empirically using 1% of scene 2534 for training over 5 trials. The values were terminal node size criterion (TNSC) 300, and impurity reduction criterion 0.001, which control tree construction. GA crossover probability 0.9, GA mutation probability 0.01, population size 60, GA selection strategy

roulette wheel, GA stopping criteria 3 generations without improvement or maximum of 20 generations. Bayesian initialization with randomization can improve the overall classification accuracy by 1.5% to 2.0% (with all other parameters being the same) and reduce the total classifier training time by over 20%.

The BTGA+ classifier was evaluated to assess how accurately the MoRAP land cover classification can be learned using only multitemporal Landsat TM spectral features. Point wise simple random sampling was used to create training data sets of different sizes. Figure 2 shows that between 1% and 5% of a single scene (about 340 thousand samples) is sufficient and that additional training data does not improve classifier performance. The slight decrease in performance beyond 30% likely is due to trees that overfit the data although the tree size stabilizes around 2800 nodes using 10% or more of the scene for training. Table 3 shows accuracy assessment of BTGA+ using a confusion table or error matrix for Scene 2534. Training used 1% of the scene and testing used the remaining 99%; 34,151,594 pixels and 33,445781 pixels for scenes 2534 and 2533 respectively. The overall accuracy is 87.3% for Scene 2534 and 83.3% for Scene 2533. The producer's accuracy are shown in the last three columns for the three datasets since we are primarily interested in the errors committed by the BTGA+.
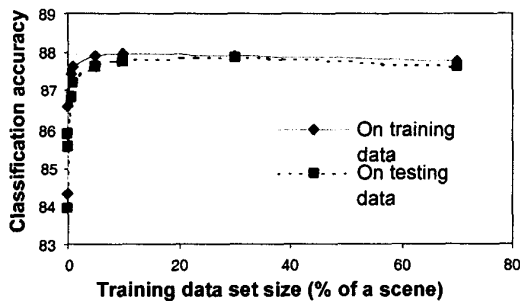

Figure 2. Influence of training dataset size on performance.

	Tree Size	Training time (sec.)	Acc. (%) training data	Acc. (%) testing data
OC1	129	1205	85.36	84.24
OC1 with pruning	84	1683	90.54	89.79
BTGA+	98	73	91.34	90.87

Table2. Overall accuracy of BTGA+ and OC1 for data AOI1.

The oblique classifier OC1 uses hill climbing to find optimal linear decision boundaries for binary tree classifiers [11]. A performance comparison between BTGA+ and OC1 was done using the AOI1 dataset. OC1 was compiled using cc -O2 on an SGI Octane (R10000, 195MHz). A 10% random sampling of AOI1 was used to train both algorithms and test on 90%. The TNSC was reduced to 30 for the BTGA+ due to the small dataset size. Table 2 shows that BTGA+ slightly outperforms OC1 even with pruning but BTGA is much faster which is important for scalability.

SUMMARY
Success of Landsat TM image pixel-based classification depends on the distinctiveness of the land cover land use spectral signatures.The BTGA+ is a new decision tree classifier that provides nearly 90% overall classification accuracy in comparison to the Missouri GAP land cover product. The enhanced BTGA+ supports Bayesian initialization of the genetic algorithm that improves classification accuracy significantly. Further improvements in classification accuracy will need to exploit ancillary data, spatial context and image segmentation.

REFERENCES
1 . The Multi-Resolution Land Characteristics (MRLC) Consortium, http://www.epa.gov/mrlc

2. Scott, J. M., M. D. Jennings, "Large area mapping of biodiversity", *Annals of the Missouri Botoanical Garden*, 85(1):34-47, 1998.

3. Scott, J.M. et al, "Gap analysis: A geographic approach to protection of biological diversity", *J. Wildlife Management*, 57(1):Supplement No. 123, 1993.

4. Drobney, R., T. Haithcoat, D. Diamond, *Missouri GAP Analysis Project*, Final Report, Missouri Cooperative Fish and Wildlife Research Unit, Dec. 1999.

5. Chai, B.B., T. Huang, X. Zhuang, Y. Zhao, and J. Sklansky, "Piecewise linear classifiers using binary tree structure and genetic algorithm", *Pattern Recognition* 29(11):1905-1917, 1996.

6. Friedl, M. A., C. E. Brodley, and A. H. Strahler. "Maximizing land cover classification accuracies produced by decision trees at continental to global scales", *IEEE Transactions on Geoscience and Remote Sensing*, 37(2):969-977, 1999.

7. Townshend, J., et al., "MODIS enhanced land cover and land cover change," MODIS land cover product algorithm theorectical basis document (version 2.0), April 1999.

8. Quinlan, J. R., *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA, 1993.

9 . Breiman, L., J. Friedman, R. Olshen and C. Stone. *Classification and Regression Trees*, Wadsworth, 1984.

10. Y. Zhu, Y. Zhao, K. Palaniappan, X. Zhou, X. Zhuang, "Optimal Bayesian classifier for land cover classification using Landsat TM data", *IEEE 2000 Int. Geoscience and Remote Sensing Symposium (IGARSS)*, Honolulu, HI, July 24-28, 2000.

11. S.K. Murthy, S. Kasif, and S. Salzberg. A system for induction of oblique decision trees. *Journal of Artificial Intelligence Research* 2(1):1-32, 1994.

12. K. Palaniappan, A.F. Hasler, M. Manyin, "Exploratory analysis of satellite data using the Interactive Image Spreadsheet (IISS) environment", *Ninth Intl. AMS Conf. Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, Anaheim, CA, Jan. 17-22, American Meteorological Society, Boston, MA, 1993, pp. 145-152.

691

Class ID	0	1	2	3	4	5	6	7	Scene 2534 Acc.(%)	Scene 2533 Acc.(%)	AOI1 Acc.(%)
0	15.91	0.37	0.00	0.28	0.00	34.59	2.53	46.32	15.91	12.79	0.00
1	3.43	6.07	0.00	0.08	0.00	1.82	0.42	88.18	6.07	79.49	83.23
2	0.06	0.05	0.00	0.02	0.00	26.55	7.14	66.18	0.00	0.00	0.00
3	0.38	0.15	0.00	86.33	0.00	7.65	0.37	5.12	86.33	91.85	95.43
4	16.96	8.00	0.00	0.00	0.00	0.74	0.37	73.92	0.00	0.00	0.00
5	0.03	0.00	0.00	0.07	0.00	90.88	3.58	5.44	90.88	91.72	95.27
6	0.22	0.01	0.00	0.12	0.00	43.01	29.26	27.39	29.26	33.95	56.68
7	0.34	0.10	0.00	0.02	0.00	3.50	1.13	94.90	94.90	87.50	95.65

Table 3. Confusion table of automatic land cover classification on scene 2534. Accuracy of scene 2533 is also shown. The training data is from 1% random sampling and the accuracy is based on testing on the remained 99% testing data.
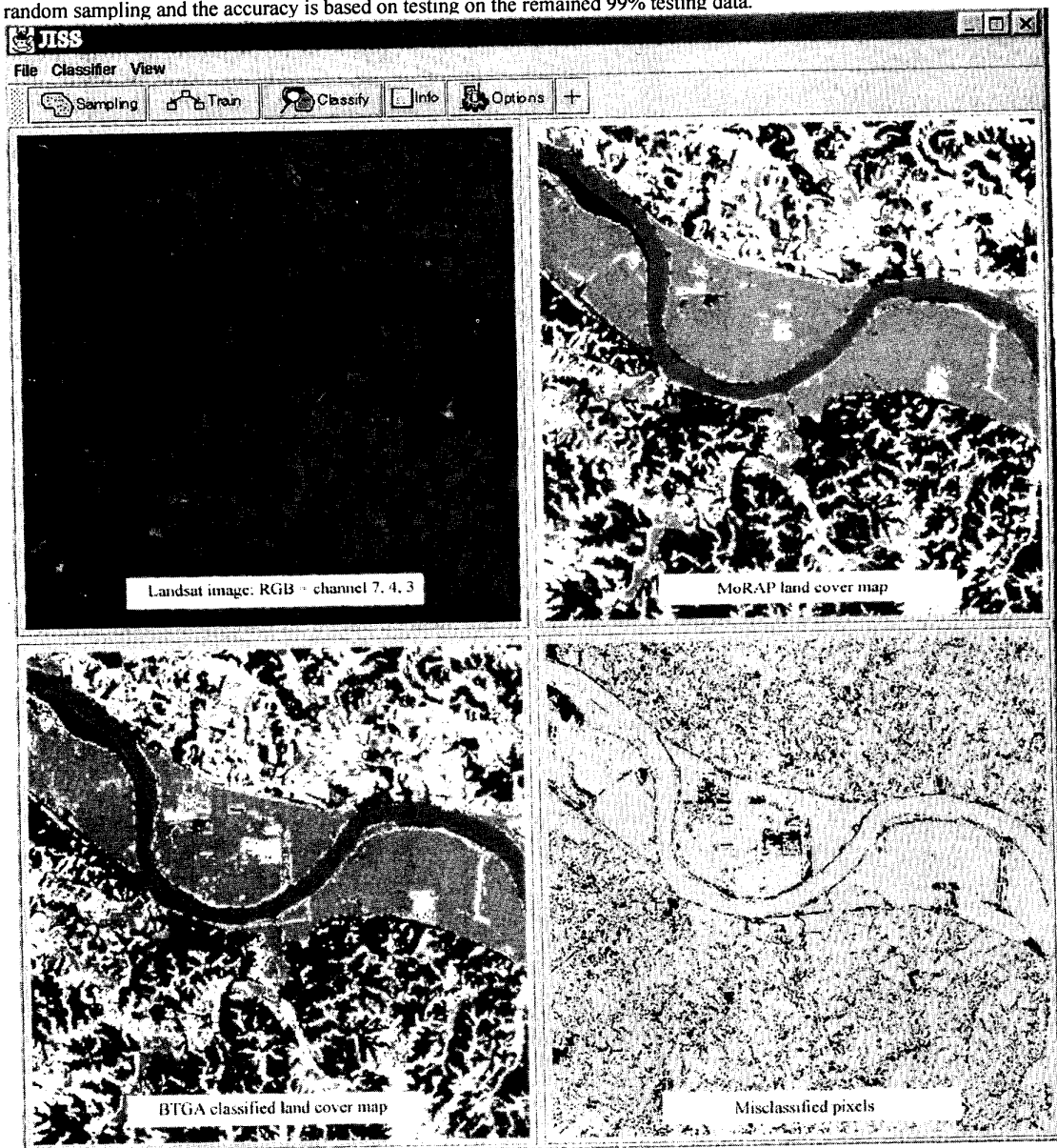

Figure 3. Java-based Image SpreadSheet tool, extending IISS [12], used to classify and visualize central Missouri land cover (AOI1) using BTGA+. The 512×512 subimage is classified with 90.9% accuracy. Cell A1 shows the Landsat data set, cell A2 shows the 8 class land cover map from MoRAP, cell B1 shows the BTGA+ automatic land cover map and cell B2 identifies the misclassified (border) pixels.