TABLE I
NUMBER OF LABELED SEGMENTS PER CLASS

Segmentation	Classes						Labeled	Total
	1	2	3	4	5	6		
watershed	788	291	156	156	57	50	1498	7708
K-means	345	243	126	67	18	30	829	6593

elongated and have $SCI > SCI_{\rm th}$, no such pattern is observed in the watershed based segmentation, indicating that most watershed-derived segments are fairly circular. This difference stems from the way the two algorithms define segments. In the watershed algorithm, segments are defined by the smallest "depressions" in the H-image, essentially by the noise in the H-image, which leads to relatively circular shapes. On the other hand, the contiguity-enhanced K-means algorithm is especially designed (see Section 3.2) to produce segments with shapes that reflect gradients of physical features. Thus, SCI is helpful in distinguishing landforms only when used in conjunction with the K-means based segmentation.

As mentioned in Section 1.3, the automatic classification of landforms is challenging because different landforms may be characterized by very similar terrain attributes, but different spatial context. For instance, segments making up craters' walls and segments constituting ridges not associated with craters may have similar values of slope, curvature, and flood, but located in different spatial contexts. Our method takes into consideration spatial context by means of neighborhood context measures. Ideally, we would like to know classes of segment's neighbors to establish its spatial context, but such information is not available prior to classification. However, even for unlabeled segments, a preliminary categorization of segments into low, medium, and high slope categories is possible on the basis of statistics of the values of $\bar{s_j}$, $j = 1, \ldots, K$. Such categorization is used to calculate the neighborhood property of a segment j, $\{a_1^s,\ a_2^s,\ a_3^s\}_j$, where $a_l^s,\ l = 1, 2, 3$ is the percentage of the segment boundary adjacent to neighbors belonging to slope category l. Similar neighborhood properties, $\{a_1^\kappa,\ a_2^\kappa,\ a_3^\kappa\}_j$, $\{a_1^f,\ a_2^f,\ a_3^f\}_j$ are calculated on the basis of curvature and flood values, yielding a total of nine features corresponding to the spatial context of a given segment. Segment-based feature vectors, denoted by u, are defined as follows when k-means based segmentation is used:

$$\begin{aligned} u \ = \ & \{\bar{s},\ \bar{\kappa},\ \bar{f},\ SCI, a_1^s,\ a_2^s,\ a_3^s,\ a_1^\kappa, \\ & a_2^\kappa,\ a_3^\kappa,\ a_1^f,\ a_2^f,\ a_3^f\} \end{aligned} \quad (12)$$

The definition is slightly modified when watershed based segmentation is used by removing the SCI component.

B. Labeled Training Set

The labeled (training) set of segments was generated by manually labeling 30% (by surface area) of the Tisia site into six classes corresponding to six landforms as described in Section 2: inter-crater plateau (class 1), crater floors (class 2), convex crater walls (class 3), concave crater walls (class 4), convex ridges (class 5), and concave ridges (class 6). Fig. 4A shows the labeled part of the Tisia site. Depending on the segmentation used, the labeled area results in a different number of labeled segments. Table I provides details on the number of labeled segments.

TABLE II
ACCURACY FOR THE THREE CLASSIFIERS AND TWO SEGMENTATIONS TECHNIQUES (STANDARD DEVIATIONS ARE GIVEN IN PARENTHESES). AN ASTERISK ON AN ENTRY REPRESENTS STATISTICALLY SIGNIFICANT IMPROVEMENT OVER THE NAIVE BAYES CLASSIFIER (AT 95% CONFIDENCE LIMITS USING A T-STUDENT DISTRIBUTION).

Dataset	Naive Bayes	Bagging with C4.5	SVM (quadratic kernel)
K-means	88.65(3.17)	$90.95(2.57)^*$	$91.06(2.50)^*$
watershed	85.27(2.51)	$90.28(1.95)^*$	$91.30(1.58)^*$

C. Classifiers

We applied different learning algorithms in our test sites for segment classification and to generate geomorphic maps. First we tried a simple classifier as a baseline for comparison to determine if a limited family of models could perform satisfactorily in our particular domain. Naive Bayes belongs to the class of generative Bayesian classifiers that estimate the posterior probability of the class label given a feature vector representing the instance (segment) using Bayes' theorem [19]. The computation of the likelihood is simplified by assuming feature independence given the class label. The independence assumption is false in our case as all the features are derived from a single source of information – the site's DEM.

Our second classifier imposes a flexible model over the feature space and was used to test the importance of reducing bias by employing a rich family of models. Support Vector Machines (SVM) is a statistical learning algorithm that works by finding an optimal hyper-plane in a (transformed) feature space [6]. The optimal hyper-plane maximizes the separation between classes. SVM exploits local data patterns and has been found to be effective in spatial data mining applications [36].

We finally employed a classifier that reduces variance through model combination. The goal was to explore the importance of reducing generalization error incurred through model variability. Bagging is an ensemble learning algorithm [7], [12]; it generates multiple models by running a single learning algorithm multiple times over bootstrapped samples of the training set. The final class label is the result of voting over the contributing models (one from each bootstrap sample). Bagging is known to work well for complex datasets and is particularly attractive when the training set is noisy [12] as is the case in our application. We use a decision tree (C4.5) as the base learner.

We employed the algorithms above as implemented in the software package WEKA [49]. We used default parameters because our empirical assessment serves simply to illustrate the potential application of this framework. Nevertheless, we performed a parameter search over the Gaussian, linear, quadratic and cubic kernels in the case of Support Vector Machines at different values of the slack variable. The best results were obtained using a quadratic kernel with the slack value ranging between $C = 1$ and 2.