MACHINE CATALOGING OF LUNAR CRATERS FROM DIGITAL TERRAIN MODEL. A. W. Bauer[1] and T. F. Stepinski[2], [1]Williams College, Williamstown, MA 01267 (awb1@williams.edu), [2]Department of Geography, University of Cincinnati, Cincinnati, OH 45221-0131 (tstepintz@uc.edu).

Introduction: An accurate and representative catalog of impact craters on the surface of the Moon is necessary to infer information about the age of surfaces and the sequence of geological events [1]. The existing global catalogs of lunar craters contain ~2700 craters [2] and ~8500 craters [3], respectively. They suffer from uneven areal coverage as they were compiled from images having different resolutions and solar illumination angles. Recently, the Lunar Orbiter Laser Altimeter (LOLA) [4], an instrument on board the Lunar Reconnaissance Orbiter (LRO) spacecraft, has acquired altimetry data which enable construction of global digital terrain model (DTM) with resolution of 64 pixels/degree. As the LOLA instrument will add new measurements even higher resolution DTM will be constructed. A DTM-based dataset is ideally suited for extraction of a global catalog of craters because it is spatially uniform and accurate. In fact, the first crater catalog of lunar crater based on the 64 pixels/degree DTM has already been constructed [5]; it contains 5185 manually identified craters with diameters ≥20 km. However, the 64 pixels/degree DTM contains many more small craters and forthcoming higher-resolution DTMs will contain even larger numbers of small craters. Catalogs incorporating smaller craters will allow finer spatial resolution of the stratigraphy. However, the only viable means to obtain such comprehensive catalogs is through automating the process of crater detection.

This contribution reports on our campaign to construct global catalogs of lunar craters from LOLA-based DTM by means of automatic crater detection algorithm. The approach follows a similar effort to auto-detect craters on the surface of Mars [6] that resulted in a global catalog of 75,919 craters listing coordinates of each crater, its diameter and depth [7]. Based on our previous experience, we expect to catalog all craters with diameters ≥3 km from the 64 pixels/degree DTM. In addition to dating surfaces, the catalog will be used for construction of global depth-to-diameter ratio maps. Here we report on our approach to constructing the catalog and on preliminary results from a limited number of sites.

Methods: The computer algorithm utilized to detect craters from DTM is based on the method described in [6]. The algorithm has two major components: a topographic depression-finding algorithm that searches for locations having potential of being craters (referred to as crater candidates) and a machine-learning-based classifier that makes a final decision whether a candidate is a crater or not. Because of computer memory considerations the global DTM is divided into overlapping tiles. The tiles overlap to accommodate craters that may be located on the edge of a tile. 108 equatorial tiles are shown in Figure 1; each equatorial tile is 24 degrees on a side and overlaps with each of its neighbors by four degrees. In addition there are 56 polar tiles not shown here. The algorithm is applied to each tile separately and the results are combined to yield the global catalog.

Depression-finding component. This component of our algorithm operates on topographic data and calculates the location and extent of all depressions present in the DTM including the nested depressions. In order to find nested depressions the algorithm performs a size-limited series of searches starting with identification of smallest depressions and ending with the largest. The final step of this portion of the algorithm is to characterize each candidate and perform preliminary elimination of candidates that are obviously not craters.

Machine-learning component. This component of our algorithm operates on the numerical data consisting of crater candidates identified by the first component. Each crater candidate is represented as a vector of five features: $\{D, d, d/D, m_2, m_3\}$. D is the candidate's diameter, d is its depth, and m_2 and m_3 are calculated values representing the elongation and lumpiness, respectively, of the candidate's shape. The role of the second component is to achieve as accurate as possible distinction between craters and non-craters. This is achieved by means of so-called supervised machine learning. A relatively small, but representative set of crater candidates is labeled by an analyst. This labeled set is submitted to an algorithm which generalizes the information contained in this set to construct a classifier – a function that will assign a label (crater or non-crater) to any other crater candidate as represented by its feature vector. In this stage of our cataloging campaign we concentrate on evaluating different possible means of constructing a classifier, looking for one that performs best on the task of crater identification.

Results: Using the 64 pixels/degree DTM we have identified crater candidates in all 108 equatorial and all 56 polar tiles. To achieve the ultimate goal of our campaign – crater catalog – crater candidates need to be classified by a machine learning algorithm. Because crater characteristics are size-dependent, we divide all candidates into three categories, small-size, medium size, and large-size, and construct a classifier for each category independently. In order to evaluate effective-

ness of different classification algorithms we selected six equatorial tiles (highlighted in red in Figure 1) showing different terrain morphologies. We have labeled 2561 small-size category candidates across these six tiles; 1271 as craters and the remaining 1290 as non craters. For our evaluation we focus exclusively on small-size candidates as this is the most challenging task; larger craters are easier to classify. The labeled craters serve both as "training" set and also as the ground truth against which the performance of a classifier is checked. Using the "training" set we built a series of classifiers to evaluate their performance. Table 1 summarizes the results.

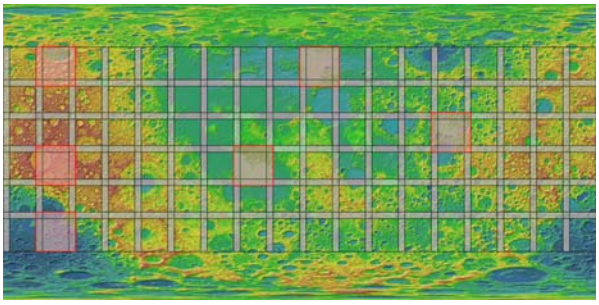

Figure 1. The tiling scheme for the equatorial region of the lunar DTM. The gray areas show the overlap between tiles. The areas highlighted in red are the locations of tiles used in our analysis.

The following classifier-building algorithms were evaluated: the C4.5 decision tree, the Random Forest, and the Support Vector Machines, all as implemented by the WEKA machine-learning software [8]. We use a concept of cross validation to assess the performance; we use training samples from five out of six tiles to construct a classifier and test the performance of such classifier on the remaining tile. This is repeated for all six tiles and the average results are reported. Two different protocols of constructing classifier were tested. In the first protocol (referred to as single) we construct a single classifier (for each algorithm) trained on the union of the labeled examples from the five tiles. This single classifier is tested on the remaining tile. In the second protocol (referred to as merged) we construct a separate classifier for each of the five contributing tiles. The testing on the remaining tile is performed by comparing the results of the five classifiers on a given candidate. Regardless of the protocol, we also have analyzed how boosting the classifier influences the results.

Conclusions: From Table 1 it is clear that the C4.5 algorithm and the merged protocol yield the best results. Thus, we plan on using this classifier to label small-size category candidates in all tiles (equatorial and polar). We still need to check whether this classifier performs equally well on medium-size and large-size candidates. Once all crater candidates in all tiles are labeled we would need to eliminate the duplicate crater finds resulting from tile overlapping. In the final phase we would perform a series of manual checks to assure a quality of the catalog. The catalog based on 64 pixels/degree DTM can be used as a training set for extraction of even more complete catalog from future higher resolution DTMs.

		Boosted?	Average accuracy	Average TPR	Average FPR
Merging	C4.5	Yes	0.712	0.935	0.473
		No	0.799	0.884	0.273
	SMO	Yes	0.661	0.920	0.575
		No	0.661	0.920	0.575
	RandomForest	Yes	0.705	0.953	0.517
		No	0.705	0.953	0.517
Single	C4.5	Yes	0.673	0.618	0.265
		No	0.676	0.623	0.263
	SMO	Yes	0.682	0.551	0.190
		No	0.682	0.551	0.190
	RandomForest	Yes	0.661	0.619	0.297
		No	0.669	0.636	0.297

Table 1. The results of machine-learning classifier analysis. The accuracy is the proportion of correctly identified craters, the TPR (truth positive rate) is the proportion of actual craters that were correctly identified, and the FPR (false positive rate) is the proportion of actual non-craters that were incorrectly identified as craters.

References:

[1] Wilhelms, D. E., The Geologic History of the Moon (U.S. Geological Survey Professional Paper no. 1348, Washington, DC, 1987).

[2] Wilhelms, D. E. et al., Proc. Lunar Sci. Conf. IX, 3735. (1978).

[3] Andersson L. B. and Whitaker B. A. (1982) *NASA Reference Publication 1097.*

[4] Smith, D. E. et al., *Space Sci. Rev.* 150, 209 (2010)

[5] Head III, J. W. et al. *Science, 329,* 1504. (2010).

[6] Stepinski T. F. et al. *Icarus, 203(1),* 77-87. (2009)

[7] Stepinski T. F. 41st LPSC #1845, (2010)

[8] http://www.cs.waikato.ac.nz/ml/weka/