A QUANTITATIVE MACHINE LEARNING-BASED MODERN TAXONOMY FOR ASTEROIDS. M. Darby Dyar[1], Sydney M. Wallace[2], Thomas H. Burbine[1], and Daniel Sheldon[3]. [1]Dept. of Astronomy, Mount Holyoke College, 50 College St., South Hadley, MA 01075 (mdyar@mtholyoke.edu), [2]Harvey Mudd College, 301 Platt Blvd., Claremont, CA 91711, [3]College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA 01003.

Introduction: Several asteroid taxonomies based on spectroscopy have been developed over the years [1-3], most recently culminating in the Bus-DeMeo (BDM) method [4,5]. But the current automated classification [6] is still challenging to use and is overdue for an update that can leverage advancements in data analysis and machine learning (ML) and include the wealth of newly acquired asteroid data. This project seeks to create a new principled classifier with quantified accuracy to replace BDM.

Data: This project uses 1,623 meteorite spectra and 694 asteroid spectra. Meteorite spectra were obtained from archives and unreleased data from the Keck/NASA Reflectance Experiment Laboratory (RELAB) at Brown University [7]. Asteroid spectra were obtained primarily from the SMASS and MITHNEOS surveys [5,8] with additional data from PDS, Polishook et al. [9], and HARTSS [10].

Methods: Bus-DeMeo classes for the first approach were determined for all 694 asteroid spectra using either the supplied labels from MIT or the online tool [11]. The second approach, based on meteorite classification [12], uses laboratory data from 27 different classes including three combinations (CV-CK, EH-EL, and IAB, IIAB, and IVB as irons). That meteorite model was then applied to the unlabeled asteroid spectra. Data were analyzed with an in-house tool written in Python utilizing the SciKit-learn library [13]. Two types of classification algorithms were tested: *Logistic regression* (LR) and *Gaussian Kernel Support Vector Machine* (SVM) [14].

Results: BDM classes were successfully reproduced using both LR (75.1% accurate) and SVM (77.7%). The best meteorite classification model used a SVM with the C and g parameters equaling 10^6 and 0.0001, respectively, after AirPLS baseline removal was applied. This gave a validation score (accuracy) of 63.8%. Another model was trained on the RELAB meteorite data without the lunar and martian meteorites, with an accuracy of 64.6%. These accuracies are likely underestimates due to false negatives for close matches.

Asteroid data were then predicted using this model to assign a meteorite label to each asteroid. Results for V-type asteroids are shown in Figure 1. These were expected to be howardites, eucrites, and diogenites, and indeed those meteorite classes constitute 89% of the objects. Another model verification was the prediction for (4) Vesta from the Dawn mission; the algorithm correctly predicted it to be a howardite. These promising results give confidence to the ML model, even in such a preliminary study utilizing only basic ML approaches. Because they are based on meteorite data, these models will ultimately provide a view of mineralogy distribution among asteroids in the Solar System.

Currently the biggest limitation on accuracy is under-representation of individual meteorite classes. ML algorithms struggle when training data are too limited because there is a larger chance that certain spectral characteristics may be missed. Work has begun to collect spectral data from recently acquired meteorites, so this drawback should rapidly be mitigated.

Future Work: These preliminary results show that ML approaches show tremendous promise for this purpose. Ultimately, results should enable direct mineralogical linkages between meteorites and their parent bodies, and provide an understanding of the distribution and abundances of objects with varying compositions throughout our solar system.

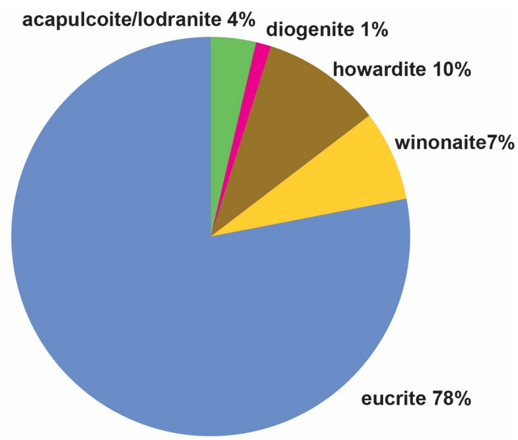

Figure 1. Percentages of V-type asteroids assigned to meteorite groups using SVM and the AirPLS baseline removal.

Acknowledgments: This project was supported by NSF grant IIS-1564083 and the RIS^{4}E node of the NASA SSERVI program (NNA14AB04A). We thank our RELAB and telescope colleagues for the use of unreleased data.

References: [1] Chapman C.R. et al. (1973) *Icarus*, *25*, 104-130. [2] Tholen D. J. (1989) Ph.D. Thesis, Univ. Ariz. [3] Bell J. F. et al. (1989) In *Asteroids II*, Univ. AZ Press, 921-944. [4] Bus S. J. and Binzel R. P. (2002) *Icarus*, *158*, 146-177. [5] DeMeo F. E. et al. (2009) *Icarus*, *202*, 160-180. [6] http://smass.mit.edu/busdemeoclass.html. [7] Pieters, C.M. and Hiroi, T. (2004) *LPS XXXV*, #1720. [8] Binzel R. P. et al. (2019) *Icarus*, *324*, 41-76. [9] Polishook D. et al. (2014) *Icarus*, *233*, 9-26. [10] Lucas M. P. et al. (2017) *Icarus*, *291*, 268-287. [11] Burbine T. H. et al. (2019) *LPS L*, Abstract #1655. [12] Wallace S. M. et al. (2020) *LPS LI*, Abstract #1036. [13] Carey C. et al. (2017) *LPSC XLVIII,* Abstract #1097. [14] Zhang Z.-M. et al. (2010) *Analyst*, *135*, 1138-1146.