APPLICATION OF MACHINE LEARNING METHODS FOR MAPPING SURFACE COMPOSITION OF THE TAURUS-LITTROW AND SURROUNDING AREA OF THE MOON. G. R. L. Kodikara and L. J. McHenry, Dept. of Geosciences, University of Wisconsin- Milwaukee, Milwaukee, WI 53211, gayantha@uwm.edu, lmchenry@uwm.edu.

Introduction: Here we demonstrate the application of advanced Machine Learning (ML) algorithms for identification and mapping of surface composition at the Apollo 17 landing site, Taurus-Littrow valley, and surrounding area of the moon. Using the Moon Mineralogy Mapper (M3) hyperspectral data, trained by the 30 reflectance spectra of soil and rock samples from the Apollo 17 mission, with six ML algorithms, we produce a surface compositional map of the area with high classification accuracy.

Method: *Spectral Analysis.* We use 14 lunar soil samples and 16 rock samples to help select the most suitable ML algorithm. The rock sample includes 4 ilmenite basalts, 6 impact melt breccias, 3 impactites, and 3 impact melt basalts [1,2]. Relevant bulk rock and bulk soil spectra were taken from the RELAB Spectral database. All spectra were resampled into M3 wavelength scale to compare directly with M3 image data. After analyzing spectral behavior for these samples, we calculated three parameters including the integrated band depths (IBD) over the 1000 nm and 2000 nm spectral regions (IBD1000 and IBD2000) and the spectral reflectance at 1578 nm (REF1578) [3]. Mafic minerals such as pyroxene and olivine have strong absorption features around 1000 nm. Pyroxene shows another absorption feature around 2000 nm as does spinel. Felsic minerals such as plagioclase and quartz don't show any absorption features in this region. Therefore, IBD tells us the presence and relative abundances of the mafic minerals [4].

We adopted six ML algorithms, including Linear Discriminant analysis (LDA), Classification and Regression Trees (CART), k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Random Forest (RF) and k-mean clustering in the R statistical software package. We used caret package developed by Max Kuhn [5] to implement the first five algorithms. We split our dataset into 10 parts, 9 to train and one for test and program runs for all combinations of train-test splits. We also repeated the same process 3 times for each algorithm to achieve the most reliable results. The k-mean clustering was also applied for the same dataset. This algorithm works iteratively to cluster each data point to one of the k groups based on their features. The number of groups is represented by the variable k. We assigned five clusters and used ward's hierarchical clustering approach to classify our dataset solely based on the values of IBD1000, IBD2000, and REF 1578.

Image Analysis. The image acquired by M3 hyperspectral camera on-board the Chandrayaan-1 mission (Image ID: M3G20090107T011405) was used for the classification after several pre-processing steps. The global linear interpolation method was applied to remove the line striping of the image. The image band at 1578 nm wavelength and IBD1000 and IBD2000 were extracted and calculated using the R statistical software package. We have selected the k-mean clustering algorithm to classify our image, after analyzing the results of ML algorithms trained by ground truth spectral data. We assigned 12 clusters for the classification to retrieve maximum information for the imaged area [6].

Results: *Spectral Analysis.* ML algorithms, LDA, CART, kNN, SVM, and RF show low mean accuracy, most likely due to the lack of ground truth data for each class (Figure 1). Accuracy tells us the percentage of observations that the model classified correctly, while the kappa statistics tell us how well two evaluators can classify an observation correctly.

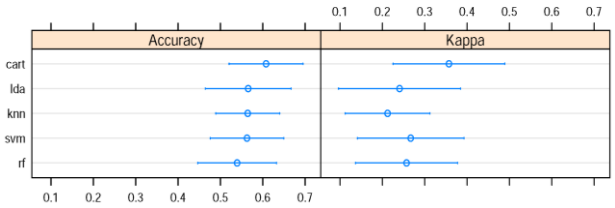

Figure 1: Accuracy and interrater reliability of adopted Machine learning methods. The different approaches yielded accuracies between 50 and 60%.

The k-mean clustering successfully classifies the given dataset with reasonable accuracies. Table 1 shows the how well the k-mean clustering method classifies the 30 rock and soil samples into 5 clusters (C1-C5). Most of the soil samples were classified into the 1st and 2nd clusters. Based on this unsupervised classification, ilmenite basalts and impactite can be clearly differentiated from the soil and impact melt basalts. Impact melt breccia shows a wide range of spectral variations due to its petrological variations.

Table 1: Classification accuracy of k-mean algorithm.

Sample	C1	C2	C3	C4	C5	Total samples in each category
Soil	5	8	0	0	1	**14**
Ilmenite Basalt	0	0	1	1	2	**4**
Impact melt Breccia	2	0	3	0	1	**6**
Impactite	0	0	0	2	1	**3**
Impact melt Basalt	1	2	0	0	0	**3**
Total Samples in each cluster	**8**	**10**	**4**	**3**	**5**	**30**

Spectral Mapping. Our mapping results [Figure 2] correctly classify the lithologies of the Taurus-Littrow valley, including the North and South Massif, the Sculptured Hills, the Lee-Lincoln scarp and even the "old" light mantle and "young" light mantle units that radiate from the northeast facing slope of the South Massif. According to our classification, Paint Splatter shows similar surface materials to Emory crater, near station 1 (Color No: 12). Stations 6 and 7 show similar materials, confirmed as impact melt breccia in the field (Color No. 10). Wessex Clef, a northeastward rising valley between the Sculptured Hills and North Massif, is compositionally distinct from both hills due to the boulders and materials (debris flow) coming from the crest of the large hills to the east [7] (Color No: 11).

Some of the fresh craters and their ejecta materials in the hills also show similar lithology as in the valley floor (Color No: 2,7,8,12). All the crests and knobs of North and South Massif and Sculptured Hills show similar surface compositions (Color No: 4 and 6), and regolith surfaces below each knob and crest show gradual variations in the materials observed, likely due to down slope mixing or space weathering (Color No: 1 and 10) [7]. We observed different surface expressions (Color No: 3) just below the Paint Splatter which is only seen in few different places over the area, especially on some crater floors.

Future Work: In the work presented here, we only use the general name of each sample for classification. We will continue to map mineral abundances using ML algorithms. The main limitation of this research was the insufficient number of spectral data to train the model. We therefore plan to incorporate the spectral and chemical data from other lunar missions to increase the accuracy of our model.

References: [1] Parker, R. A. et al. (1973) Apollo 17 Preliminary Science Report, NASA SP-330, 707pp. [2] Meyer, C. (2008) Lunar Sample Compendium. [3] Cheek, L. C. et al. (2013) Journal of Geophysical Research, 118, 1805-1820. [4] Pieters, C. M. and Englert, P. A. J. (1993) Remote geochemical Analysis: Elemental and mineralogical composition. 594 pp. [5] Kuhn, M. (2008) Journal of Statistical Software, 28(5), 26 pp. [6] Harrington, P. (2012) Machine Learning in Action, 354 pp. [7] Schmitt, H. H. et al. (2017) Icarus, 298. 2-33.

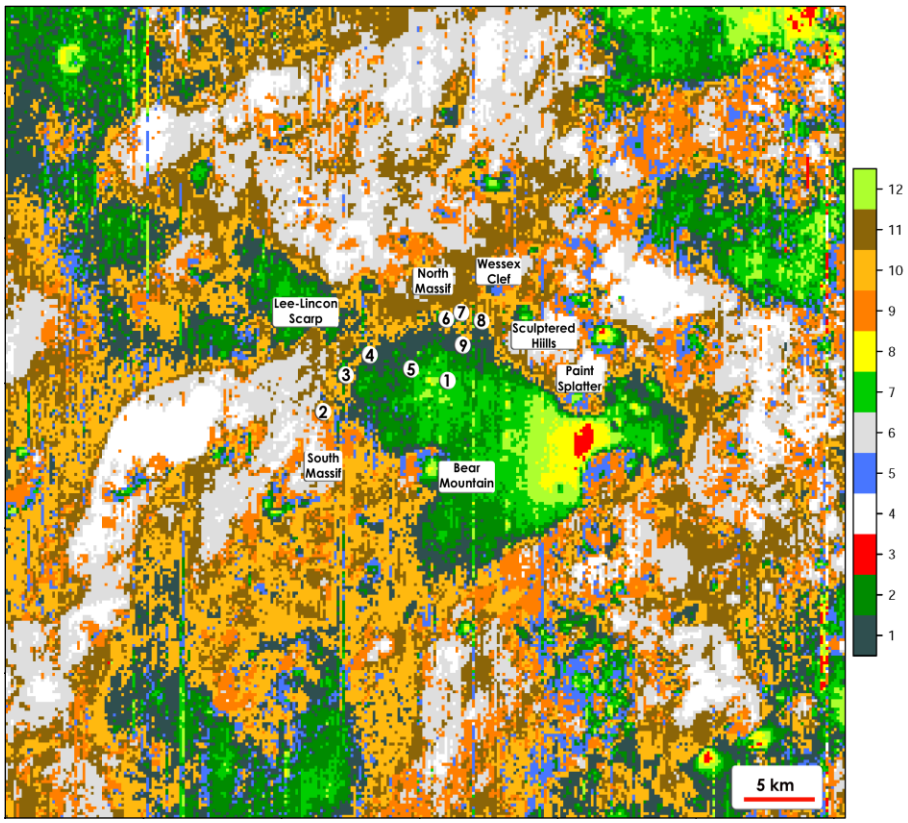

Figure 2: Classified Surface Composition Map for the Taurus-Littrow valley, based on our spectral mapping results. Colors indicate 12 different compositional classification groups. Named features and stations associated with the Apollo 17 landing site and surrounding area are labeled.