IDENTIFYING CHEMICALLY DISTINCT CLASSES OF LUNAR ROCKS USING DIMENSIONALITY REDUCTION ALGORITHMS. S. Shubham[1], B. Cohen[2], and R. Arevalo Jr[1] [1]University of Maryland, College Park (sourabh@umd.edu MD, USA 20742), [2]NASA Goddard Spaceflight Center (MD, USA - 20771).

Introduction: Instrument techniques such as Laser Induced Breakdown Spectrometry (LIBS) or Gamma Ray Spectrometry (GRS) have the capability of quantifying the elemental composition of rock and regolith. GRS and LIBS have been a part of previous planetary investigations such as Lunar Prospector (LP) and Mars Science Laboratory (MSL), we well as future instruments and mission concepts such as Bulk Elemental Composition Analyzer (BECA) [1] and Potassium Argon Laser Experiment (KArLE) [2] respectively. The quantification of elemental abundances in the rocks can help in understanding the petrogenesis of different rock types. Therefore, an unambiguous identification of chemically distinct rock types on the Moon is important to support the findings in future lunar missions.

In this study, we use unsupervised machine learning algorithms – Principal Component Analysis (PCA), and t-distributed stochastic neighbor embedding (t-SNE) – to classify chemically distinct groups of lunar rocks using major oxide data. By performing PCA and t-SNE, we identify clusters of lunar rocks that are compositionally similar and test the validity of the classification using a subset of compositional data of lunar samples. A simple plot between elemental ratios such as Mg# (=Mg/(Mg+Fe)) and elemental quantities, e.g. FeO (wt%) have traditionally been used to classify lunar rocks due to distinctive chemical makeup [3]. Here, we compare the accuracy of classifying lunar rock by PCA or t-SNE methods, vs the simplistic 2D plots of Mg# vs FeO (wt%) to test for unambiguous identification of lunar rock classes.

Dataset: We consider the Gruithuisen Domes (GD) region of the Moon as a test case scenario to identify the expected rock types and their expected composition at a local scale. GD is a region expected to contain abundant silicic rocks such as granites and rhyolites [4], and has been selected by NASA as the landing site for the PRISM 2 landed payload, set to deliver the Lunar VISE mission suite in 2027.

Expected rock classes. The rocks at GD have been interpreted to be dominated by felsic rocks (possibly granites), anorthosites (Ferroan Anorthosite (FAN)), Mg-suite rocks (Norite), mare basalts, and KREEP basalts (impact melts). Using LP-GRS data, we determine the TiO_2 (wt%) and FeO (wt%) of the rocks at GD to identify suitable geochemical analog for each classes from the Apollo samples. We use criteria such as <2% TiO_2 for silicic rocks and 8-10 wt% FeO for norites. Identifying basaltic rocks is straightforward due to their distinctive dark color in color imagery. However, among the other rock types – namely anorthosites, norites, and granites – we test PCA and t-SNE to find a robust classification technique.

Train/Test datasets. We compile a database using major element data for each of the three rock classes – FANs, norites, and granites. For FANs and norites, we use MoonDB (now a part of Astromaterials Data System) [5] to compile a training dataset. Due to the limited number of granitic clasts in returned lunar samples, we compile a training database for granites using terrestrial granites data that are accessible through the GEOROC database [7] and test the validity of the classification using Apollo granite data. We build a testing dataset using the major elements of Apollo sample 60025 as the analog for ferroan anorthosite; sample 14318 for norites, and sample 12013 for granites. The entire compilation has 44 datapoints corresponding to each class (43 for FAN) for computing the principal components and t-SNE dimensions.

Method: We use PCA algorithm, which is a dimensionality reduction algorithm often used for better visualization of data as clusters (e.g. [6]). PCA utilizes a linear transformation of a dataset into a new coordinate system such that the first few columns contain maximum variance. A 2-D or 3-D plot between the first to third principal components are often used to visualize clusters among datasets. Comparatively, t-SNE projects the dataset into a new coordinate system by performing non-linear transformation of the data. We recognize the clusters visually and compare the number of points of a class that can be visually and unambiguously identified within a cluster to test for accuracy in classification. We also perform the prediction of test data within these clusters.

Results: *PCA Results*. Out of 44 datapoints, 47.7% of norite data fall within non-overlapping norite cluster. overlap with anorthosite cluster, 84.1% anorthosite datapoints fall within non-overlapping anorthosite clusters, and 97.7% granite datapoints fall in granite clusters. The test data corresponding to the Apollo samples 12013 (granite), 14318 (norite), and 60025 (FAN) are all correctly recognizable among their respective clusters (Fig. 1).

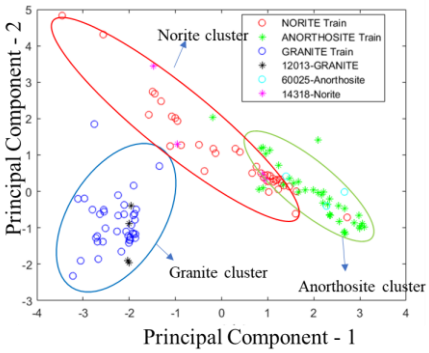

Figure 1. PCA on major oxide data*. Apollo Samples 12013 (representative of lunar granite), 60025 (FAN), and 14318 (Norite) major oxides are tested by projeting in PC biplot.*

t-SNE Results. Using t-SNE, we recognize more distinctive clusters for each rock type (fig. 2) as compared to PCA. 84.1% of norite data fall under non-overlapping norite clusters (as compared to 50% in PCA), 93.18% anorthosite data fall under anorthosite clusters, and 95.45 % granite data fall under granite cluster. The limitation with t-SNE process is the non-repeatability of t-SNE results with each analysis (as the process is heuristic); however, local convergence among each cluster remains same (i.e., non-overlapping data will follow the same trend in each iteration).

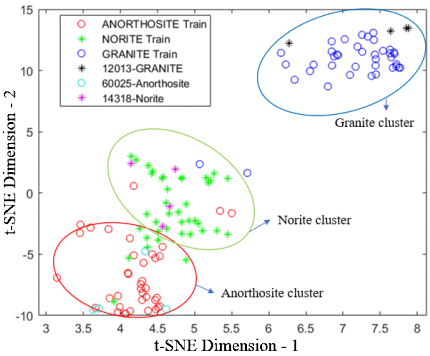

Figure 2. t-SNE on the major oxide data *(same dataset as those used in figure 1). t-SNE performs a non-linear transformation of geochemical data to perform clustering among similar classes. Compared to PCA results as shown in fig. 1, t-SNE proves to be better in classifying the three classes.*

2-D plots between chemical indices. Mg# is often used as a common geochemical classifier for identifying magnesian suite rocks such as norite. Anorthosites and norites have relatively higher values of Mg#, while anorthosites have lower FeO (wt%) as compared to norites. So, 2-D plot between Mg# and FeO (wt%) can be used to distinguish different types of lunar rocks (fig. 3). However, the test data for FAN, granite, and norite from Apollo samples are unrecognizable due to overlapping clusters. None of the Apollo sample 12013 (granites) and Apollo sample 60025 (FAN) could be unambiguously recognizable within the clusters.

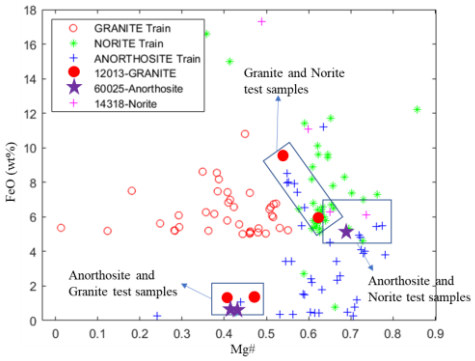

Figure 3. 2-D plot between Mg# and FeO (wt%). *Ferroan Anorthosites are clustered at values with higher Mg# than granites and comparable Mg# like Norite. Ferroan anorthosites have lower FeO (wt%) than norites. However, there is ambiguity in identifying Apollo 12013 granite from the anorthosite clusters and norite clusters. Additionally, 2 data points of FAN Apollo sample 60025 could be misrecognized as norites due to comparable FeO (wt%)*

Conclusions: In the test-case scenario for chemically classifying the analogous non-mare rock types of GD (i.e., FAN, norite, and granite), we identify t-SNE as the most robust clustering algorithm as compared to PCA method, while the traditionally used 2-D plots between Mg# vs FeO (wt%) was least robust. Furthermore, in both PCA and t-SNE, the test-data are unambiguously recognized with 100% accuracy, although it should be noted that only 4 data points corresponding to each analogous Apollo samples were used here. We also notice that a 2-D plot between Mg# and FeO (wt%) can be used to create separate clusters of the chemically distinct lithologies, however the FANs and granites test data are non-predictable as they do not plot over any unambiguous clusters.

Acknowledgments: S. Shubham is supported by grant 80NSSC19K1398. Data analysis and plotting were performed in MATLAB programming language. We thank Dr. Dan Moriarty III for interpreting Gruithuisen Domes lithology using remote sensing tools. All data used were accessed from MoonDB (now a part of AstroMAT).

References:

[1] Parsons, A. M. et al. (2016) *(NSS/MIC/RTSD). IEEE, 1-3.,* [2] Cohen, B. M. et al. (2019) *Astrobiology 19.11 (2019): 1303-1314.* [3] Wang X., and Zhao S. (2017) *JGR Planets 122*(10), .2034-2052. (1996) *LPS XXVII*, 1344–1345. [4] Ivanov, M. A. et al. (2016) *Icarus, 273, 262-283.* [5] Lehnert, K., et al. (2018), *PSIDAC Vol. 2082. No. 2081*. [6] Cone K.A. et al., (2020*) Icarus, 346, 113787*. [7] GEOROC https://georoc.eu/georoc/new-start.asp