UTILITY OF GENERATIVE APPROACHES FOR METEORITE AND ASTEROID CLASSIFICATION.
Sydney M. Wallace[1], M. Darby Dyar[1,2], Thomas H. Burbine[2], and Daniel Sheldon[3]. [1]Planetary Science Institute, 1700 East Fort Lowell, Suite 106, Tucson, AZ 85719 (smwallace@g.hmc.edu); [2]Dept. of Astronomy, Mount Holyoke College, 50 College St., South Hadley, MA 01075; [3]College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA 01003.

Introduction: Asteroids date back to the creation of the Solar System and can provide valuable insights into its composition and formation. With advancements in remote spectroscopy, huge improvements have been made in classifying asteroids to better our understanding of them. The Bus-DeMeo (BDM) taxonomy [1] leverages slope scores of reflectance spectra and uses principal component analysis to cluster similar samples together. However, the current BDM automated classification program does not make use of newly acquired asteroid data or the advancements made in machine learning (ML). This project combines these new developments and provides new findings about the composition and creation of the Solar System.

There are few physical asteroid sample returns we can leverage as a ground truth for the classification of asteroids, though more samples are soon forthcoming (Hayabusa 2 [6] and OSIRIS-Rex [7]). As proxies for asteroids, meteorites are smaller pieces of asteroid that have fallen to Earth; their compositions, petrology, and mineralogy have been thoroughly studied. Thus, our understanding of the spectroscopic characteristics of meteorites can provide a ground truth for models built to classify asteroids. How well the model handles asteroid-specific challenges must be highly prioritized during model selection.

The task of asteroid classification includes many issues that need to be carefully considered when building and selecting a model. Relatively few meteorite and asteroid spectra are available compared to the millions of examples other commonly used ML datasets can supply (images, pages of text, etc.). Instrument-dependent noise also affects spectral measurements. This problem is less relevant when building models for meteorites, for which most data have been acquired at one facility (e.g. Keck/NASA Reflectance Experiment Laboratory (RELAB) at Brown University [3]). But given the small number of asteroid spectra available, data from many different telescopes must be used. Therefore, classification models that can handle spectra taken across multiple instruments are required.

This paper highlights the importance of these factors in model selection. Preliminary results are reported on the use of Fisher Discriminant Analysis (FDA). Findings from [2] are extended here to include meteorite spectra from two spectrometers rather than just one.

Data: This project uses two separate sub-datasets from our collection of meteorite spectra. The first dataset (we will denote as RELAB-only) has 1,621 meteorite spectra spanning 27 classes. This set only includes spectra (from archives and unreleased) from RELAB and was resampled to 0.01 μm resolution over a wavelength range of 0.3 to 2.5 μm. The second dataset (we will denote as Mixed) includes 156 spectra from both RELAB and the University of Winnipeg (PSF) [8]. In total, there are 68 L-type, 56 H-type, and 32 eucrite spectra that have been acquired at either facility. These spectra were resampled to 0.01 μm resolution over a wavelength range of 0.35 to 2.5 μm.

Methods: We used in-house tools built in Python from the Scikit-Learn library [4] and from Kernel FDA (KFDA) [5], which is a library that is compatible with Scikit-Learn. In [2], a *Support Vector Machine* (SVM) with a *Gaussian kernel* performed the best out of all the classification models tested. FDA is again implemented here as the first generative approach to be tested.

Generative models assume that there is a functional probability over all of the classes while discriminative models assume there is a functional probability of a class when given a spectrum. Because generative approaches model all of the possible classes, they are better at handling cases when the test data may lie outside of the training data distribution and when there are relatively small amounts of training data. For these reasons, a generative model was selected for performance comparison to the discriminative models analyzed in [2].

KFDA is a specific type of linear discriminant analysis (LDA), which is a dimensionality reduction technique closely related to principal component analysis. It works by defining new axes that maximize the distance between the means of each class and minimizing the variation within each class. The kernel defines how this “distance” metric is defined. To have a more direct comparison between discriminative and generative approaches, we used the same kernel (Gaussian) as the SVM.

Results of KFDA trained on RELAB-only Meteorites: A Gaussian Kernel FDA with a $\gamma = 0.075$ achieved a 59% accuracy on the validation set, only ~6% worse than the best discriminative model. As seen in Figure 1, many of the missed predictions are

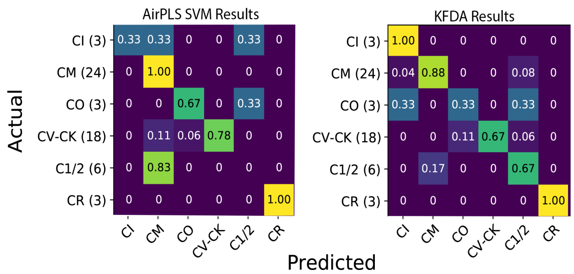

Figure 1. The confusion matrices for the carbonaceous chondrites in the validation set that were correctly and incorrectly classified. Each square indicates the proportional value that was identified for each class. 1.0 implies 100% accuracy.

misclassified as meteorites that are closely related to the true label. Although this model did not outperform previous models across all 27 classes, it performs equally as well on the carbonaceous chondrite classes (CI, CM, CO, CV-CK, C1/2, and CR) as highlighted in Figure 1. Both the SVM and the KFDA classified 44/57 of these type of meteorites correctly. However, KFDA did better at correctly classifying more of the meteorites that had fewer examples. This can be clearly seen with how the C1/2 meteorites are classified by both models. The SVM classifies most examples as CM, the carbonaceous chondrite class that has the most examples. KFDA classified most of these meteorites correctly, showing it learned the class itself rather than classifying those meteorites as the most likely class (CM). This shows that KFDA is generally better at learning classes with few examples.

Results of KFDA and SVM on the Mixed set of meteorite spectra: The top-performing SVM model without baseline removal had the same hyperparameters as the top performing model from the previous section. These hyperparameters did not fare well with this experiment because SVM models struggle when only a few examples of each class are available for training. Thus, we selected hyperparameters based on the performance score from the training data. Three separate models were trained on three different sets of training data: RELAB data, PSF data, and a 50/50 mix of data. The results shown in Table 1 display the performance score for each of the three models trained on the three training sets and their scores on the three test sets.

The KFDA hyperparameters performed well on the smaller datasets with no tuning required other than reducing the maximum number of components used, because the number of classes was reduced from 27 to 3, and this value can go no higher than one less the number of classes in the dataset. As described earlier, the structure of a KFDA model allows for the model to perform better when handling classes with only a few examples.

SVM

Trained on:	**RELAB**	**PSF**	**Mixed**
RELAB	0.611	0.500	0.611
PSF	0.611	0.694	0.611
Mixed	0.611	0.528	0.582

KFDA

Trained on:	**RELAB**	**PSF**	**Mixed**
RELAB	0.806	0.583	0.667
PSF	0.583	0.806	0.667
Mixed	0.667	0.586	0.806

Table 1. The top and bottom tables show results for the SVM and KFDA models respectively. The far-left column indicates what kind of data the model was trained on (RELAB, PSF, or Mixed) and the following columns show what testing data were used for the score.

Discussion: Even if the KFDA model did not perform as well as the SVM with the RELAB-only dataset, it was more promising in its ability to detect patterns in the less-populated classes. Furthermore, when we tested both models against smaller datasets that had measurements from two different instruments, KFDA performed better than SVM, except for one scenario when the model is trained on PSF data and tested with RELAB data. This further proves that KFDA can better detect class-specific patterns when data examples are limited and can better handle the challenges of instrument noise.

Summary: There are important challenges to consider in selecting the best model for asteroid classification. There are fewer than a thousand total spectra and they are all taken from various instruments around the world. Luckily, we can leverage data from meteorites to better understand this relationship and to also supply a ground truth to classification labels. Generative approaches, such as KFDA, can better perform under these task-specific challenges.

Acknowledgments: Supported by NASA 80NSSC19M0215 to the RISE[2] SSERVI.

References: [1] DeMeo F. E. et al. (2009) *Icarus, 202,* 160–180. [2] Wallace S. M. et al. (2020) *LPS LI,* Abstract #1036. [3] Pieters C.M. & Hiroi T. (2004) *LPS XXXV,* Abstract #1720. [4] Pedregosa et al. (2011) *JLMR 12,* 2825-2830. [5] Ghojogh B. et al. (2019) *arXiv:1906.09436*. [6] Yoshikawa M. et al. (2012) *Meteorites, Comets, Meteors,* Abstract #6188. [7] Lauretta D. S. et al. (2012) *LPS XLIII,* Abstract #2491. [8] https://ctape.uwinnipeg.ca/FACILITIES/.