January 2019

Applying Machine Learning To Asteroid Classification Utilizing Spectroscopically Derived Spectrophotometry

Kathleen Jacinda Mcintyre

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://commons.und.edu/theses

APPLYING MACHINE LEARNING TO ASTEROID CLASSIFICATION
UTILIZING SPECTROSCOPICALLY DERIVED SPECTROPHOTOMETRY

by

Kathleen Jacinda McIntyre
Bachelor of Science, University of Florida, 2011
Bachelor of Arts, University of Florida, 2011

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

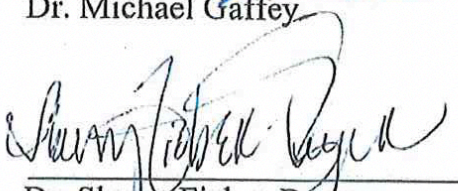in partial fulfillment of the requirements

for the degree of

Master of Science

Grand Forks, North Dakota

May
2019
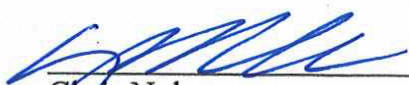
This thesis, submitted by Kathleen Jacinda McIntyre in partial fulfillment of the requirements for the Degree of Master of Science from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Dr. Michael Gaffey

Dr. Sherry Fieber-Beyer

Dr. Taufique Mahmood

This thesis is being submitted by the appointed advisory committee as having met all the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

Chris Nelson
Associate Dean of the School of Graduate Studies

5/2/19
Date

PERMISSION

Title   Applying Machine Learning to Asteroid Classification
                  Utilizing Spectroscopically Derived Spectrophotometry

Department   Space Studies

Degree   Master of Science

        In presenting this thesis in partial fulfillment of the requirements for a
graduate degree from the University of North Dakota, I agree that the library of
this University shall make it freely available for inspection. I further agree that
permission for extensive copying for scholarly purposes may be granted by the
professor who supervised my thesis work, or in his absence, by the Chairperson
of the department of the dean of the School of Graduate Studies. It is understood
that any copying or publication or other use of this thesis or part thereof for
financial gain shall not be allowed without my written permission. It is also
understood that due recognition shall be given to me and to the University of
North Dakota in any scholarly use which may be made of any material in my
thesis.

                                          Kathleen Jacinda McIntyre
                                          April 23, 2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to my advisory, Dr. Mike Gaffey, and the members of my committee (Dr. Sherry Fieber-Beyer, Dr. Nels Forsman, Dr. David Whalen, and Dr. Taufique Mahmood) for their guidance and support throughout my time in the master's program at the University of North Dakota. I would also like to thank my parents, John and Susan McIntyre, who put up with more than a lot over the years but never stopped believing in me. And a special thanks to my sister, Tina McIntyre, who ensured my survival with coffee and granola bars while I was writing this thesis. I couldn't have done it without all of you.

ABSTRACT

Taxonomic studies of asteroids have been ongoing for more than fifty years without a clear understanding of the class parameters. The current method of Principal Component Analysis is computationally expensive and leaves ambiguous results. In this study, I selected the machine learning algorithm, k-Nearest Neighbor in combination with the current Bus-DeMeo (DeMeo, et al. 2009) taxonomic classification schema to test if machine learning can take the place of Principal Component Analysis. Using a dataset of spectrophotometric color indices derived from combined visible and near-infrared (NIR) observations and paired with Bus-DeMeo taxonomic class, I created a training dataset for the model to learn. The results support the visible wavelength region as more diagnostic of spectral slope and the NIR wavelength region as more diagnostic for surface mineralogy. The overall accuracy scores (>80%) of the machine learning test dataset validate the methodology, but fall short of the threshold necessary to replace current methods of classification (>95%). The overall robustness of the Bus-DeMeo taxonomy is corroborated through the relatively similar grouping structure between the C-, S-, and X-complexes in both wavelength ranges, suggesting an overall relationship between slope and qualities present across multiple wavelength regimes. This is possibly due to spectral features being closely tied to surface mineralogy and spectral reddening of the slope believed to be tied to the effects of space weathering.

1 MOTIVATION

1.1 Why Asteroids?

There are currently over a half million numbered and catalogued asteroids in our solar system. A little over two hundred years ago, there were none. It was not until 1801 when Giuseppe Piazzi was searching for the elusive planet theorized to be between Mars and Jupiter (Kepler 1596; Bode 1772), that he discovered (1) *Ceres*. In the following years (2) *Pallas*, (3) *Juno*, and (4) *Vesta* were all discovered around the same orbital distance. The gap between Mars and Jupiter had been filled, not by a single large planet, but seemingly by multiple smaller bodies. The theory was soon developed that these smaller bodies were the broken pieces of a fragmented planet (Olbers 1803). It was such a captivating proposal that it took over a century for a new idea to gain traction. In the mid-1900s, Otto Schmidt proposed that asteroids were not the result of a catastrophic collision, but rather a failure of a planet to fully aggregate. The current understanding of asteroid formation is that they are the original building blocks of our solar system and studying them offers us a window back in time to the early days of our planet's formation.

The theory of solar system formation began when a large cloud of molecular gas and dust became massive enough to trigger gravitational collapse. The majority of the mass in the cloud went to the center, where the Sun formed, and the rest flattened into a protoplanetary disk (Mizuno 1980; Pollack 1984; Boss 1997, 1998). It was from this disk that the planets and the asteroids formed (Chambers & Wetherill, 1998). The early inner solar system was mostly rocky material, while the outer solar system was cold enough for volatiles and ices to condense on to aggregating bodies. This led to the primary divide in the main asteroid belt, where

siliceous S-type asteroids dominant the inner region of the main asteroid belt and carbonaceous C-type asteroids dominate the outer areas of the Main Belt (Fisher 1941, Kitamura 1959, Wood & Kuiper 1963, Chapman et al. 1971, Chapman et al. 1975).

The two panels on Figure 1-1 represent the asteroid population of the inner solar system. The majority of the asteroids are in the Main Belt, located between Mars and Jupiter, (1.52-5.20 AU), but there are also significant asteroid populations around the orbit of Jupiter (Trojans) and near Earth's orbit as well (near-Earth asteroids, or NEAs). Compare the area around Earth's orbit in the left panel of Figure 1-1 to that in the right panel of Figure 1-1. The NEA population is sparse compared to the Main Belt, but there are thousands of catalogued bodies for us to investigate in our cosmic backyard.
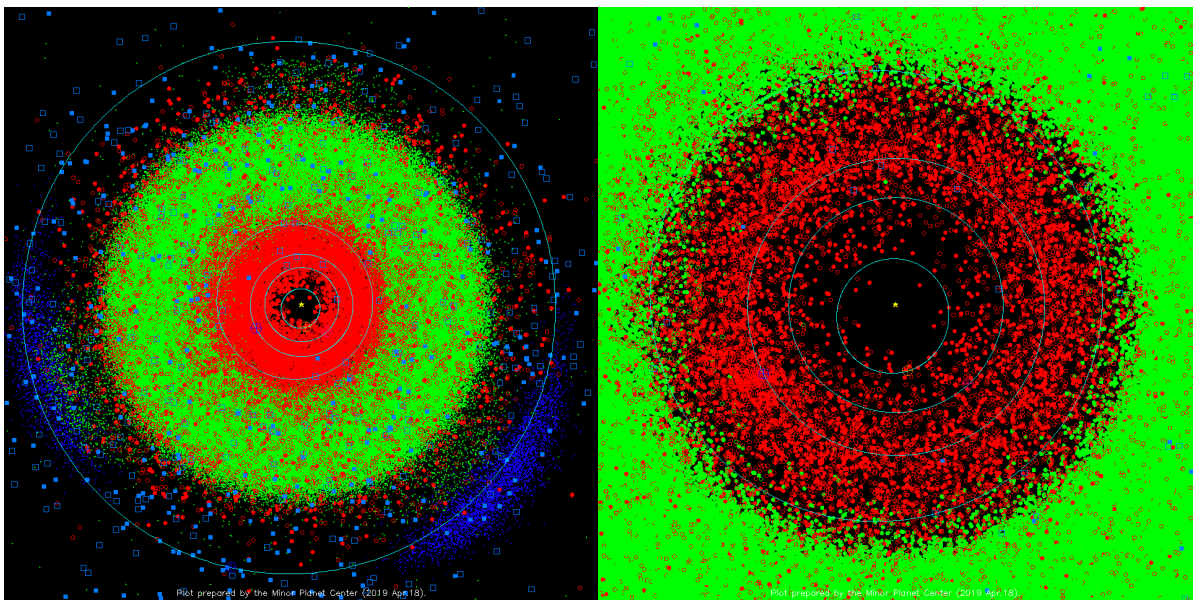
Figure 1-1: Plots showing the known asteroids and comets, as of April 18, 2019[1]. Green circles represent Main Belt asteroids, red circles represent NEAs, blue circles are Jupiter Trojans, and comets are shown as blue squares. The Sun is represented by the yellow star in the center. The left panel depicts the solar system out to the orbit of Jupiter, with the orbits of the five innermost planets shown. The right panel provides a better view of the NEA population, extending out to the orbit of Mars.

---

[1] https://cgi.minorplanetcenter.net/iau/lists/MPLists.html

Asteroids are of great scientific value as they allow scientists to study the early solar system, but there are more reasons to pay attention to asteroids. The general definition of an asteroid is a rocky small solar-system body that orbits the Sun at heliocentric distances ranging from interior to Earth's orbit to Jupiter's orbit (IAU 2006). While the general asteroid distribution of volatile rich outer-belt asteroids and rocky, thermally processed inner-belt asteroids gives us clues as to the conditions of the early solar system, there is still much we do not understand about how the orbital properties of asteroids evolved over time. Assigning asteroid families is one way to try to determine dynamical origins and migrations of asteroids. Asteroid families are believed to be the product of the collisional breakup of a progenitor body (Hirayama 1918). This breakup results in many smaller asteroids, which can be traced back to their parent body through orbital dynamics. To further assign an asteroid as a member of a family, its spectra can be used to genetically link members together (Farinella et al. 1992). A homogeneous progenitor would result in compositionally similar family members, but if the progenitor was a differentiated body, there should be family members of varying compositions and taxonomic classifications, making family classification more complex.

Of great importance to these dynamical studies is trying to understand and predict how asteroids migrate inwards, potentially towards Earth. NEAs are asteroids with perihelion distances less than or equal to 1.3 AU and aphelion distances greater or equal to 0.983 AU. In an effort to associate an impact risk factor with orbital properties, NEAs have been subdivided into four classes: Atiras are NEAs whose orbits are entirely encompassed within the orbit of Earth. Atens are NEAs with Earth-crossing orbits but a semi-major axis less than Earth's. Apollos are NEAs with Earth-crossing orbits that have semi-major axes greater than Earth's. Finally, Amors are NEAs whose orbits are outside of Earth's but within 1.3AU of the Sun at perihelion. Asteroids whose Minimum Orbit Intersection Distance with Earth is equal to or

less than 0.05 AU are known as Potentially Hazardous Asteroids (PHAs). The IAU currently

lists 724 NEAs with non-negligible probabilities of impacting Earth (IAU 2019). One such

PHA is Bennu, currently being visited by the OSIRIS-REx spacecraft.
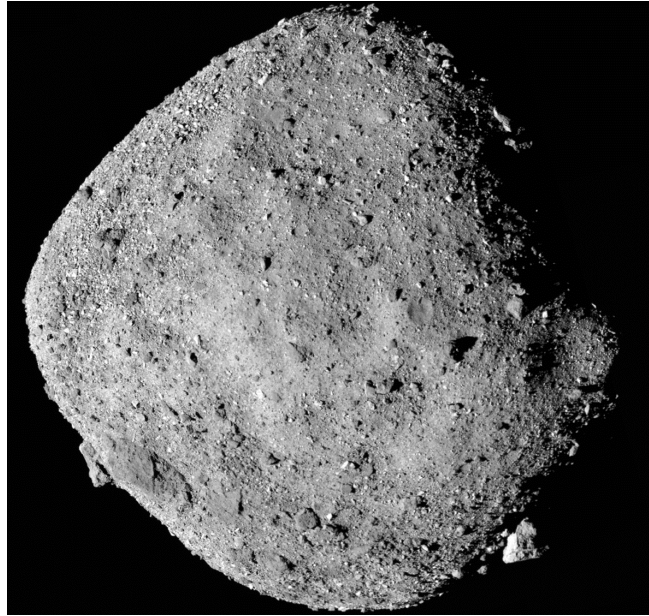
Figure 1-2: Compilation of eight images of asteroid Bennu taken with OSIRIS-REx's
PolyCam instrument on 12/02/2018 (NASA/Goddard/University of Arizona).

What makes NEAs a threat to Earth is also what makes them prime for scientific study:

proximity. Asteroids do not emit significant amounts of their own light in the visible and near-

infrared (NIR), so spectroscopic and photometric studies use the reflected light from the Sun,

which I will discuss in Section 2. They also tend to be faint and difficult to observe. NEAs are

closer than the majority of asteroids located in the Main Belt, and therefore relatively brighter.

They are also the topic of the new field of space resource allocation (Fieber-Beyer & Gaffey

2019). Asteroids formed from the same protoplanetary disk as Earth, and so the majority have

a low relative inclination (Lecar & Franklin 1973). In addition to a low inclination, NEAs are

close to Earth's orbit, making them the easiest targets to visit for scientific and resource

utilization (Yarnoz et al. 2013). Motivating this study is the desire to better understand the properties of these asteroids through remote sensing and modern computational tools.

1.2 Statement of Purpose

The primary objective of this paper is to further the understanding of the relationship between asteroid classifications and observational properties by combining historical tools and perspectives with the computational capabilities of the present. To meet this objective, I will apply a machine learning algorithm to derived spectrophotometric data and compare to current classification efforts through the Bus-DeMeo taxonomy (DeMeo, et al. 2009).

1.3 Outline

Section 2 will examine the relevant background and historical information necessary to support the motivation behind this work. Section 2 outlines why scientists strive to classify observations into groups, so it only follows that Section 3 explores the techniques commonly used to justify these classifications. Section 4 defines the observations and describes the steps followed to implement a machine learning routine using derived spectrophotometry. Finally, Sections 5 and 6 discuss the results, conclusions, and future work that can utilize this method.

2   BACKGROUND

2.1 Reflectance Spectroscopy

Reflectance spectroscopy is an effective measurement for the physical characterization of asteroids. Reflectance spectroscopy studies light that has been reflected, refracted, or scattered from a material, as a function of wavelength. For example, a mineral on the surface of an asteroid is hit by incident photons. Some photons will be absorbed by the mineral, some will be transmitted through the grains, while others will be scattered via reflection and refraction. What we are detecting when we observe that asteroid are the scattered photons, leaving an absorption band feature in the spectra where the absorbed photons are missing from the continuum (Burns 1970).

The absorbed photons are converted to heat and re-emitted as thermal radiation (McCord & Adams 1977), which will affect the flux received from the asteroid. Also, the Sun's peak emission (~0.5 μm) is located within the wavelength region included in this study (0.8-2.5 μm), so reflectance will be maximized. Additionally, Earth's atmosphere is relatively transparent in the visible band, with telluric corrections applied to the NIR band, making the visible-NIR the ideal wavelength region to study asteroid spectroscopy.

2.2 Spectral-Altering Surface Properties

There are multiple factors contributing to the alteration of an asteroid's spectrum: chemical composition, space weathering, particle size, temperature, and albedo (Burns 1970; Adams 1974; Adams & McCord 1971; Hiroi et al. 1994; Lebofsky et al. 1986). This is why asteroid

surface characterization and diagnostic taxonomic classifications are so difficult using remote sensing techniques. We can only speculate the surface mineralogy of asteroids through observational spectra alone.

2.2.1 Mineralogical Composition

The most diagnostic wavelength region for mineralogy of silicate-based compounds is in the NIR spanning the $0.8 - 2.5$ μm range (across the 1 and 2 μm absorption bands of olivine and pyroxene). Studies of the spectral properties of minerals in the NIR region have concentrated primarily on olivine and pyroxene (Clark 1957; Burns 1970; Adams 1974; King & Ridley 1987; Cloutis & Gaffey 1991). Minerals are characterized by a unique reflectance spectrum with different spectral features. The features specific to a certain mineral come from the electronic and vibrational transitions within the molecules. At what wavelengths these features will appear in the spectrum depends on the chemical species involved and the structure of the mineral (Burns 1970).

Olivine and pyroxene appear to be the two most abundant minerals on the surface of asteroids, but the possibility that you can obtain a similar spectrum through the combination of various minerals cannot be ignored. The primary diagnostic features of olivine is comprised of three overlapping individual bands located in the ~1 μm region, while the primary diagnostic features in pyroxenes are two distinct absorption bands located near 0.9 μm and 2 μm (Clark 1957; Burns 1970). Section 3 will discuss how to use these band features to distinguish the mineralogical subtypes of the S-class asteroids. To that end, Gaffey et al. (1993) developed a classification scheme based on the ratios of olivine to pyroxene in the spectra. An advantage of exploring mineralogical sub-types is that the effects of space weathering (see section 2.2.2) are virtually nonexistent, making the classification system sensitive to only composition (Gaffey et al. 2010). Figure 2-1 depicts the seven mineralogical subtypes along the olivine-

orthopyroxene mixing line from olivine-dominated at S(I)-type to pyroxene-dominated at S(VII)-type.
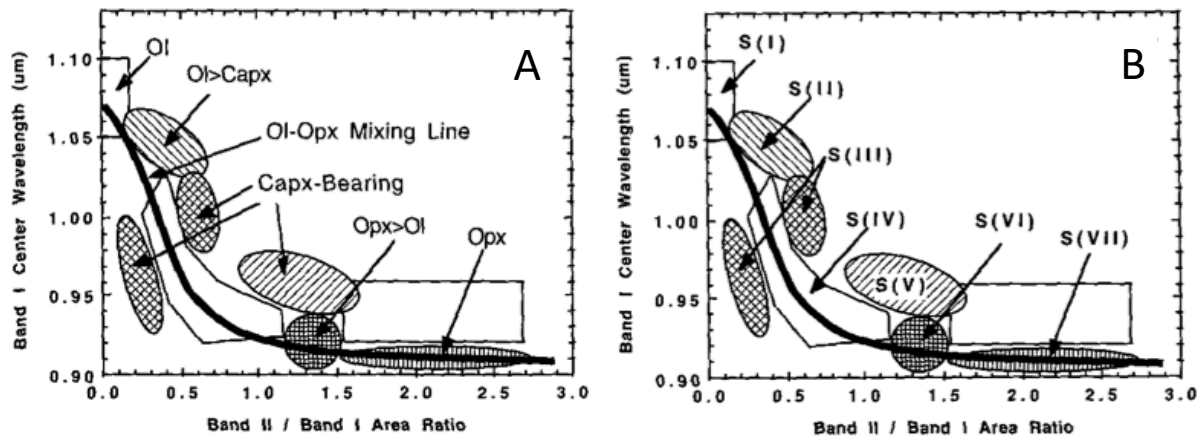
Figure 2-1: Plots of the Band I center vs. the Band II/Band I area ratio. (A) General representation of the mineralogical zones of the S-type subgroups. (B) The corresponding subgroups labelled S(I) – S(VII). Adapted from Figure 1E and 1F of Gaffey et al. 1993.

2.2.2 Space Weathering

An essential step in our comprehension of asteroids requires the understanding of the processes resulting in space weathering. All airless bodies in the solar system are affected by space weathering, including asteroids. Effects of space weathering are induced through solar ion irradiation, including solar wind, solar flares, and coronal mass ejections, galactic cosmic-ray ion irradiation, and micrometeoroid bombardment (Clark & Johnson 1996; Moroz et al. 1996; Hapke, 2001; Sasaki et al. 2001). Figure 2-2 in displays these processes and the resulting sputtering, displacement, and vaporization of asteroidal surface particles, which we observe as space weathering effects.

Space weathering was first predicted to occur on the lunar surface (Gold, 1955), and return samples from the Apollo program confirmed that space weathering of the lunar regolith caused its spectra to exhibit weaker features and a redder continuum (Adams & McCord 1971; Pieters et al. 2000; Noble et al. 2001; Taylor et al. 2001). Extrapolating this theory to asteroids,

it becomes apparent that space weathering may alter the appearance of remote sensing observations and can explain the missing link between ordinary chondrite meteorites and their S-type asteroid parent bodies, which plagued planetary scientists for decades (Bell et al. 1989).
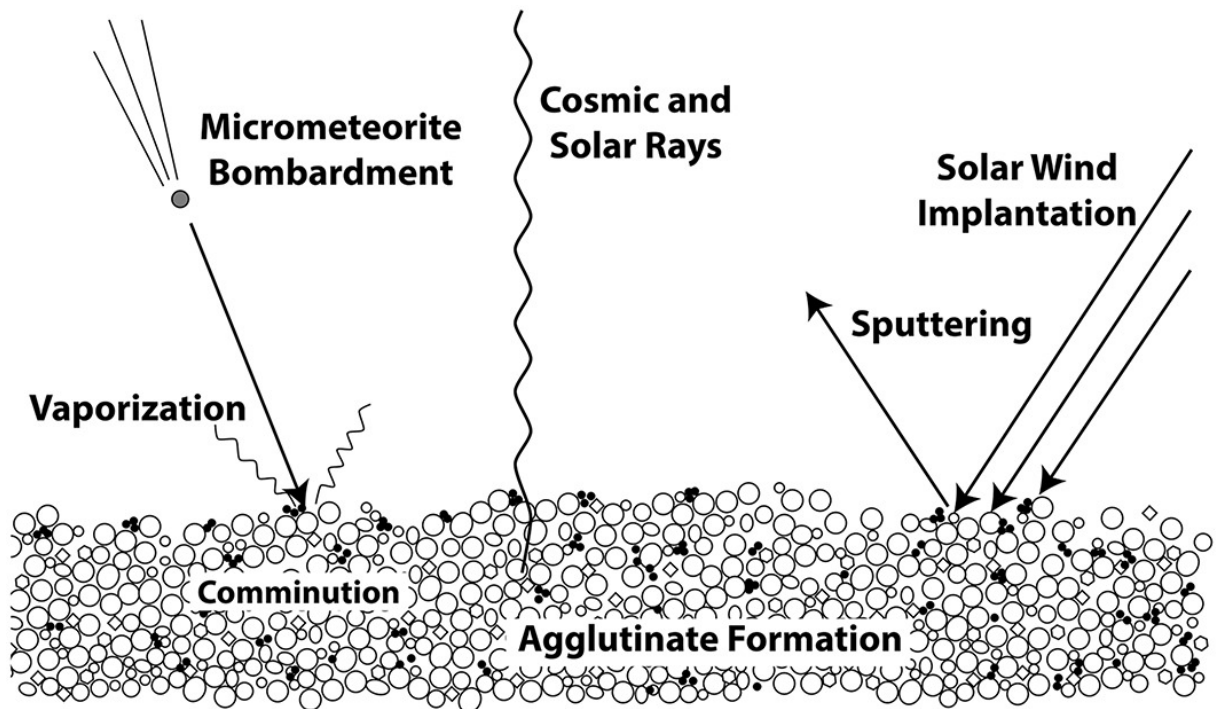
Figure 2-2: Depiction of the general space weathering processes occurring on asteroidal surfaces. Adapted from Figure 1 of Pieters & Noble 2016.

Irradiation from solar wind or galactic cosmic rays can alter the surface grains through high-energy interactions that result in sputtering or chemical alteration (Keller & McKay 1997). Solar wind particles and galactic cosmic rays both primarily consist of hydrogen ions. The flux of solar wind impacts is much higher than the galactic cosmic ray flux, but the galactic cosmic rays are much more energetic and so cannot be discounted (Brunetto et al., 2015). Brunetto and Strazzulla (2005) describes an experiment where samples of silicate-rich rock were irradiated with various ions (H^+, He^+, Ar^+, Ar^{2+}) at energies ranging from 60 to 400 keV,

to simulate space weathering on asteroids. The reflectance spectra were both reddened and darkened in the 0.25 – 2.7 μm range, but Brunetto and Strazzulla believe that while ion irradiation may dominate in the inner solar system, micrometeoroid bombardment could be the main contributor to space weathering in the asteroid belt.

2.2.3 Particle Size

The particle size distribution of the surface regolith of asteroids can also impact its reflectance spectrum. Figure 2-3 compares the reflectance spectra between Vesta and a Howardite EET87503 of different grain sizes, ranging from less than 25 μm to 125 μm (Hiroi et al. 1994).
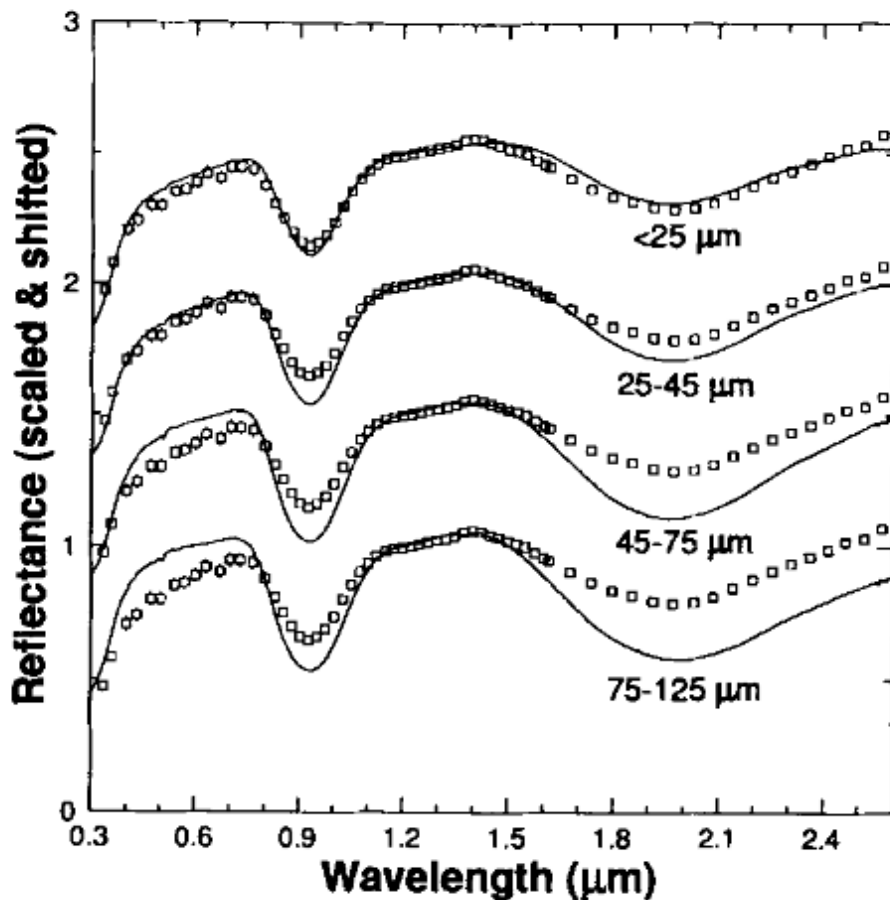
Figure 2-3: Vesta's normalized reflectance spectra compared to a Howardite EET85703 of varying particle size across the visible and NIR. Figure 2 from Hiroi et al. 1994.

We can see that fine particles fit the spectrum of Vesta well, indicating that the surface of Vesta is covered with a fine regolith. Also of note is how the spectrum changes as the grain size increases. The overall slope becomes bluer and the absorption band depths increase. Grain size does not affect the NIR as much as the mid-IR because the mid-IR region is dominated by thermal emission, which is more strongly affected by changes in surface grain structure (Hargrove, 2009). The Dawn spacecraft gave us a close-up view of Vesta and its regolith (Russell et al. 2012, 2013). Material excavated from large craters on Vesta indicates that the regolith is more than 1-kilometer-thick in some places (Denevi et al. 2016), supporting the hypothesis of a thick, fluffy layer of regolith fines.

2.2.4 Temperature and Albedo

Surface temperatures of asteroids can vary wildly depending on heliocentric distance. As noted in Section 1, asteroids are defined as small bodies from within Earth's orbit to Jupiter's orbit. This means the surface temperature of airless bodies at these semi-major axes ranges from about 120 K for the Trojans to over 300 K for near-Earth asteroids (Hinrichs et al., 1999).

The surface temperature of asteroids is also linked to albedo (Lebofsky et al. 1986):

$$T = \sqrt[4]{\frac{(1-A)L}{\eta\varepsilon\sigma 4\pi r^2}} \tag{2-1}$$

where T is the equilibrium surface temperature, A is the bolometric Bond albedo, L is the luminosity of the Sun, η is the "beaming parameter" describing rotational and thermal inertia properties, ε is the emissivity of the asteroid in the infrared, and r is the asteroid's heliocentric distance.

Temperature can affect the spectra of asteroids because the depth and width of absorption bands of olivine and pyroxene can change at different temperatures (Roush, 1984; Singer & Roush, 1985; Burbine et al. 2009), and therefore, to a lesser degree, with different albedos. Changes in the shape of these absorption features can affect the mineralogical interpretation of the asteroid spectra (Lucey et al., 1998; Moroz et al., 2000). Work done by Burbine et al. (2009), Reddy et al. (2012), Sanchez et al. (2012) and Sanchez et al. (2014) have developed correction factors for temperature changes, as well as for effects introduced by changes in phase angle, for meteorite analogs of the V-type, S(IV)-type, and A-type asteroids, but more work is needed on meteorite analogs with dominant mineralogies other than olivine or pyroxene.

2.3 Taxonomy

Asteroid taxonomy is a means by which to identify statistically significant similarities (or differences) amongst a collection of asteroid spectra. It should be noted that taxonomy is not diagnostic of surface mineralogy of asteroids. Asteroid classes are based on observational spectral information only. While it is likely that asteroids of different spectral classes will be found to be compositionally different, the assumption that members of the same taxonomic group will be found to be similar is not valid. Over the last five decades, more than ten asteroid taxonomies have been suggested (Chapman, Morrison & Zellner 1975; Bowell et al. 1978; Gradie & Tedesco 1982; Tholen 1984; Barucci et al. 1987; Tedesco et al. 1989; Gaffey et al. 1993; Howell et al. 1994; Birlan et al. 1996; Bus & Binzel 2002ab; DeMeo 2009), focusing on spectral properties such as color, albedo, absorption bands, and slope in UV, visible and, later, NIR regions to classify asteroids. Each successive taxonomy can be seen as a progression from

previous efforts, and as higher-resolution spectroscopic data becomes available, this effort continues (Usui et al. 2019).

2.3.1 Early Classifications

The first asteroid taxonomy was comprised of only two classes, C and S, with unusual spectra being categorized under U (Chapman, Morrison, and Zellner 1975). The link between the classifications and the S-class's absorption features were similar to olivine and pyroxene, and the C-class's relatively flat, featureless spectra were similar to that of the carbonaceous chondrites. Chapman, Morrison, and Zellner (1975) stated that the classes were not based on inferred surface compositions, but rather on the observed spectral features of the U-B and B-V colors.

David Tholen created a commonly used taxonomy based on color and albedo (Tholen, 1984). Tholen generated 14 classes, the most abundant of which were the C-class asteroids and the S-class asteroids. Other classes included the A, B, D, F, G, and T based on spectral features, and the additional E, M, and P classes that were featureless (Tholen, 1984). The remaining three classes: Q, R, and V were created each for a specific asteroid that could not be classified into one of the other eleven. Tholen's (and subsequently Bus's and Bus-DeMeo's) taxonomy is based on principal components of their spectra, which groups the asteroids with similar spectral features, not composition. Tholen's taxonomy was not challenged until the arrival of charge-coupled devices (CCDs) and the Small Main-Belt Asteroid Spectroscopic Surveys, SMASS (Xu et al. 1995; Xu et al. 1996) and SMASSII (Bus & Binzel 2002a; Bus & Binzel 2003). These surveys resulted in a visible spectra database three times the size of that used by Tholen.

The Bus taxonomy (Bus & Binzel 2002ab) was the first major asteroid taxonomic overhaul since Gaffey et al. (1993) introduced the concept of using a mineralogical

classification scheme. Unfortunately for Bus and Binzel, given the constraint on wavelength region from 0.44 – 0.92 μm, they were unable to probe any of the NIR absorption features. They used a Principal Component Analysis to try and maintain consistency with Tholen. Their first principal component was a linear fit to the overall slope and the second principal component was intended to be sensitive to the presence of a 1 μm olivine or pyroxene feature, but it is unclear if there is any compositional diagnostics in their classification or if it is primarily due to space weathering. Figure 2-4 shows the boundary regions for the three largest complexes, where slope represents the first principal component.
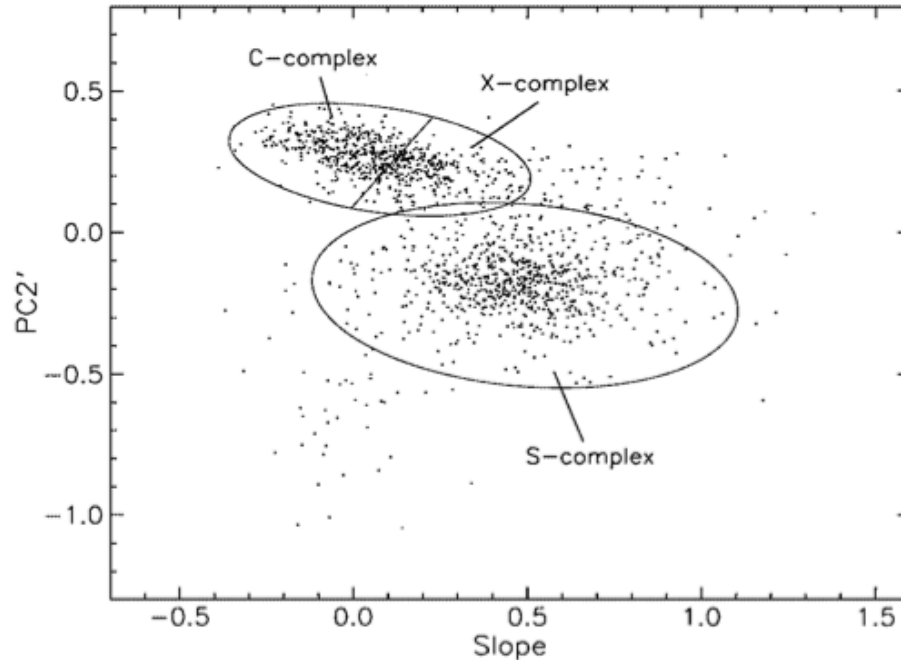
Figure 2-4: Principal Component Analysis plot showing the boundaries for the three main Bus taxonomic complexes. Figure 1 from Bus & Binzel 2002a.

2.3.2 Bus-DeMeo Taxonomy

The Bus-DeMeo (DeMeo et al., 2009) taxonomy expands on the Bus taxonomy by using 371 reflectance spectra in the NIR. The same general techniques were used, but the slope was removed from the spectra prior to the Principal Component Analysis in an attempt to

diagnostically extract information from the spectral features. Section 3 will discuss Principal Component Analysis in more detail. Figure 2-5 shows examples of the average spectrum of each of the 24 taxonomic classes, highlighting the sub-classes of the three main complexes, and Table 2-1 lists the classes and sub-classes with a short description on each. In this paper, I have assigned Bus-DeMeo classifications to the observations in my analysis and utilize them as the known outcomes for training and testing the machine learning algorithm.

Figure 2-5: Average visible and NIR spectra of each of the 24 Bus-DeMeo taxonomic classes. Adapted from Figure 15 of DeMeo et al. 2009.

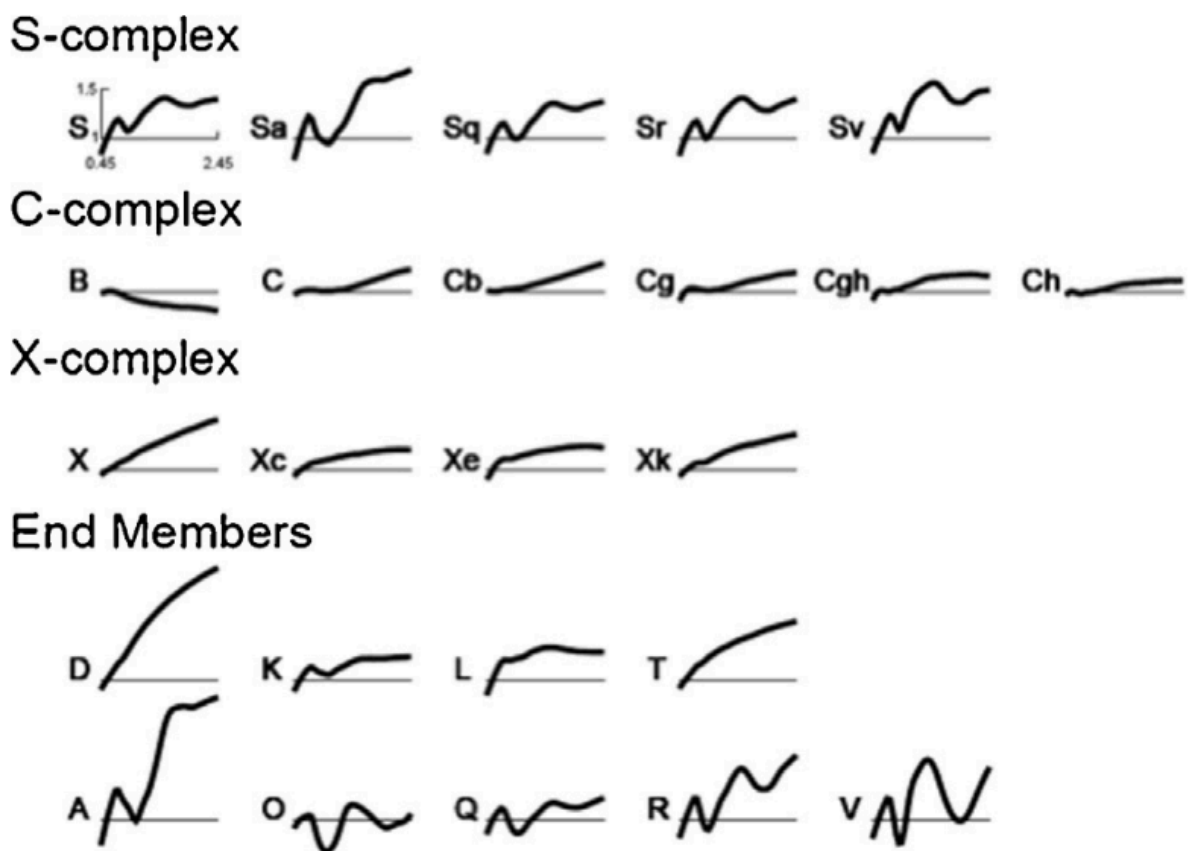| Class	Description
A	Deep and extremely broad absorption band with a minimum near 1 μm, may or may not have shallow 2-μm absorption band; very highly sloped
B	Linear, negatively sloping often with a slight round bump around 0.6 μm and/or a slightly concave up curvature in the 1- to 2-μm region.
C	Linear, neutral visible slope often a slight rough bump around 0.6 μm and low but positive slope after 1.3. May exhibit slight feature longward of 1 μm.
Cb	Linear with a small positive slope that begins around 1.1 μm.
Cg	Small positive slope that begins around 1.3 μm and pronounced UV dropoff.
Cgh	Small positive slope that begins around 1 μm and pronounced UV dropoff similar to Cg also includes a broad, shallow absorption band centered near 0.7 μm similar to Ch.
Ch	Small positive slope that begins around 1.1 μm and slightly pronounced UV dropoff also includes a broad, shallow absorption band centered near 0.7 μm.
D	Linear with very steep slope, some show slight curvature or gentle kink around 1.5 μm.
K	Wide absorption band centered just longward of 1 μm, the left maximum and the minimum are sharply pointed and the walls of the absorption are linear with very little curvature.
L	Steep slope in visible region leveling out abruptly around 0.7 μm. There is often a gentle concave down curvature in the infrared with a maximum around 1.5 μm. There may or may not be a 2-μm absorption feature.
O	Very rounded and deep, "bowl" shape absorption feature at 1 μm as well as a significant absorption feature at 2 μm.
Q	Distinct 1-μm absorption feature with evidence of another feature near 1.3 μm; a 2-μm feature exists with varying depths between objects.
R	Deep 1- and 2-μm features; the 1-μm feature is much narrower than a Q-type, but slightly broader than a V-type.
S	Moderate 1- and 2-μm features. The 2-μm feature may vary in depth between objects.
Sa	Has a deep and extremely broad absorption band at 1 μm; has similar features to A-types but is less red.
Sq	Has a wide 1-μm absorption band with evidence of a feature near 1.3 μm like the Q-type, except the 1-μm feature is more shallow for the Sq.
Sr	Has a fairly narrow 1-μm feature similar to but more shallow than an R-type as well as a 2-μm feature.
Sv	Has a very narrow 1-μm absorption band similar to but more shallow than a V-type as well as a 2-μm feature.
T	Linear with moderate to high slope and often gently concaving down.
V	Very strong and very narrow 1-μm absorption and as well as a strong 2-μm absorption feature.
X	Linear with medium to high slope.
Xc	Low to medium slope and slightly curved and concave downward.
Xe	Low to medium slope similar to either Xc- or Xk-type, but with an absorption band feature shortward of 0.55 μm.
Xk	Slightly curved and concave downward similar to Xc-type but with a faint feature between 0.8 to 1 μm.

Table 2-1: Descriptions of the spectral classes and sub-classes in the Bus-DeMeo taxonomy. Adapted from Table 5 of DeMeo et al. 2009.

3   SPECTRAL ANALYSIS TECHNIQUES

3.1 Principal Component Analysis

Principal Component Analysis is an analytical tool used to reduce the dimensions of a dataset in a way that retains most of the information in the original dataset through coordinate transformations. The difficulty of classifying asteroids into similar groups is a problem with high dimensionality when each wavelength becomes a dimension. Principal Component Analysis has been a popular method of determining asteroid taxonomy (Tholen 1984; Bus & Binzel 2002b; DeMeo et al. 2009). While the goal of asteroid taxonomy is to group asteroids in ways that are meaningful to their properties, Principal Component Analysis does not distinguish the asteroid's compositional information. Instead, the goal of Principal Component Analysis is to reduce the number of variables in such a way that retains interpretable information through transforming a group of correlated variables into another group of uncorrelated variables; the resulting variables are labeled as Principal Components.

Tholen (1984) was the first to apply Principal Component analysis to asteroid taxonomy. He used the transformation equation:

$$PC = E(F - B) \tag{3-1}$$

where PC is the principal component, E is the vector of the covariance matrix of color indices, and (F – B) represent the normalized color indices. Tholen reduced the dimensionality to just two principal components, PC1 and PC2, correlated to the depths of the 0.3 and 1.1 μm features (U-V and V-X colors), respectively, as seen in Figure 3-1. This figure shows the fourteen classes Tholen derived from ~500 asteroid observations using the Eight Color Asteroid Survey
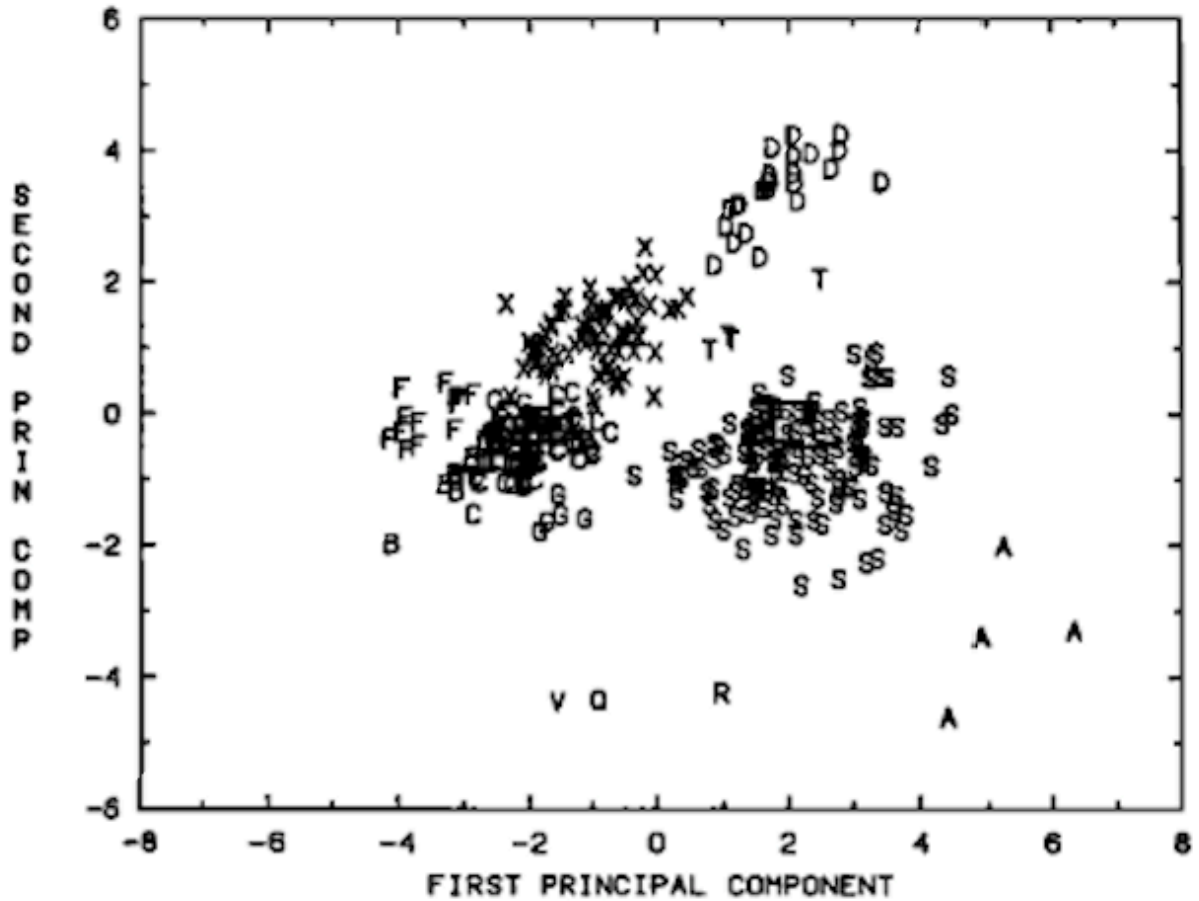
Figure 3-1: PCA projection of the 7-dimensional tree cluster structure that Tholen used to derive his taxonomy. Figure 8 from Tholen 1984.

(ECAS) (Zellner et al. 1985). The axes on Principal Component Analysis plots do not have physical meanings, but it is generally understood that the horizontal axis represents the first Principal Component. In this plot we can see clear groupings of the S-, C-, X-, and D-types that are still seen in taxonomic studies today.

The Principal Component Analysis method employed by DeMeo et al. (2009) began by removing the overall slope of the spectrum prior to further analysis. Bus and Binzel (2002b) determined that the slope was the first principle component in their taxonomy. DeMeo et al. (2009) tested this and found that the slope accounted for 88.4% of all the variance in the data.

DeMeo wanted the Principal Component Analysis to be sensitive to other features. Prior to the Principal Component Analysis, DeMeo also utilized a mean centering method, where the average value in each wavelength channel was subtracted from the dataset, leaving each channel with a mean value of zero. The wavelength channel at 0.55 μm was also removed, as that was the value normalized to unity, reducing dimensionality further prior to the analysis. DeMeo (2009, Equation 2) transposed the mean-subtracted data set through:

$$PC_x = [E_x^T][D^T]$$

(3-2)

where PC_x is the principal component of index x, E_x is eigenvector of index x, and D is the reflectance spectral values of each asteroid. This is fundamentally similar to the method used in Tholen (1984).

3.2 Band Parameters

NIR reflectance spectra of S-type asteroids contain two absorption features near 1 and 2 μm that are diagnostic of mineralogy (Gaffey et al. 1993). Band I (BI) refers to the 1 μm feature and Band II (BII) refers to the 2 μm feature. The presence of iron-bearing silicate minerals is what causes the BI and BII absorptions (Burns et al. 1972; Adams 1974). In Figure 3-2, the spectra of olivine and low-Ca pyroxene (Clark et al. 2007) are shown, normalized and off-set so that the combined olivine and pyroxene 1 μm features that make up BI can be seen. The tangential lines running over the maximum reflectance are used to determine the band areas for the Band Area Ratio (BAR).

Cloutis et al. (1986) proposed a mineralogical method to probe these absorption bands by taking the ratio of the orthopyroxene over the sum of the orthopyroxene and olivine. To calculate this, we need to find the ratio of BII to BI. We can see in Figure 3-2 that we would
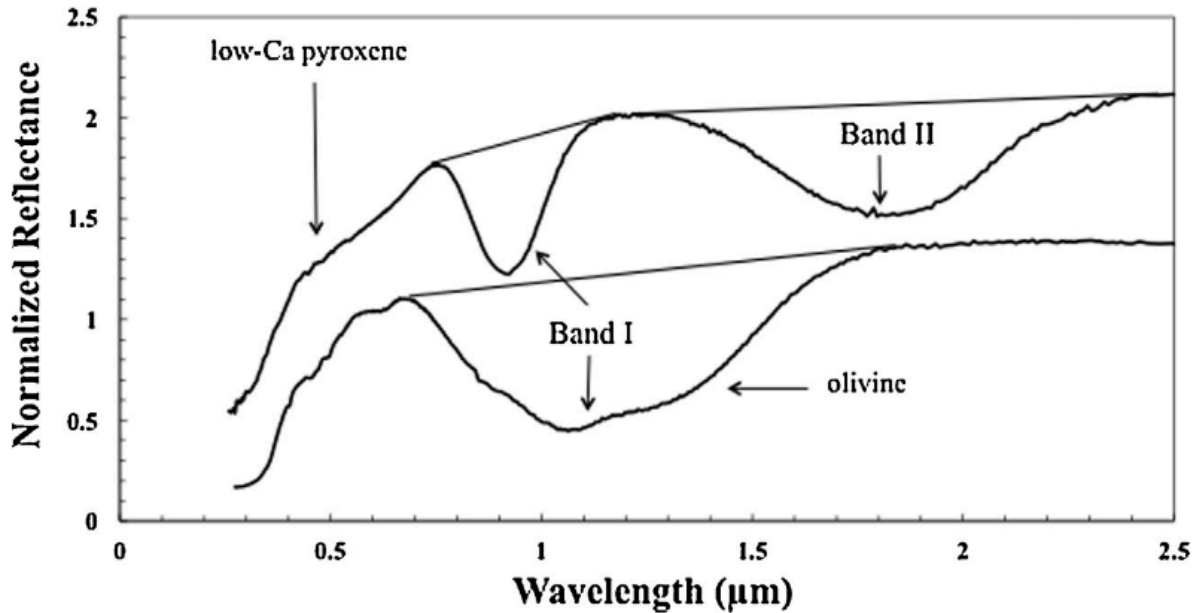
Figure 3-2: Plot of the offset reflectance spectra of olivine and low-Ca pyroxene with BI and BII areas shown. Figure 3 from Burbine 2016.

expect a maximum to be located ~0.7 μm, a minimum to be located ~1 μm, another maximum to be located ~1.3-1.7 μm, and another minimum ~2 μm. A tangential line is then fit from the first maximum to the second maximum and from the second maximum to the end of the spectra. Caution should be used when approaching the red edge of the spectra. The sensitivity at 2.50 μm may not be acceptable and a shorter wavelength cutoff should be considered. The spectrum is divided by the two tangential lines, removing the continuum. Now all the spectral information that remains pertains to the absorption features. The area calculated under these tangential lines gives the BI area and BII area, respectively. The ratio of BII area to BI area is known as the Band Area Ratio (BAR) (Cloutis et al. 1986). The BAR can be used to calculate the ratio of the orthopyroxene over the sum of the orthopyroxene and olivine using the Equation 1 from Fornasier et al. (2003):

$$\frac{OPX}{OPX+OL} = 0.4187 \times (BAR + 0.125) \qquad (3\text{-}3)$$

where OPX stands for orthopyroxene and OL is olivine.

3.3 Machine Learning

Machine learning, at its core, is another form of statistical analysis we can apply to asteroid classification, with an emphasis on the use of computational power. The benefit of machine learning over other statistical methods is that it can predict outcomes of new datasets based on its learning algorithm. The drawback is that the model requires training to learn; however, once a machine learning algorithm has completed its training, it can be made publicly available for others to benefit from this time-consuming training.

Machine learning algorithms are made up of a development algorithm, a minimization function, a model, and a dataset. This concept is best summarized by Mitchell (1997): "A computer program is said to learn from experience E with respect to some class of tasks T and perform measure P, if its performance at tasks in T, as measured by P, improves with experience, E." In the case of asteroid classification, the task (T) is classifying, so the machine learning algorithm contains a training dataset of previously classified asteroids (experience E), and if it has successfully learned, it will accurately predict future classifications (performance measure, P). The application of this concept is described in Section 4.

The idea of applying a machine learning model to asteroid spectra is not a new one. In the early 1990s there were several papers devoted to using low resolution spectroscopy for classification purposes (Ninomiya & Sato, 1990; Hepner et al. 1990; Howell et al., 1994; Merenyi et al., 1996). Howell et al. (1994) used Artificial Neural Networks to accurately classify the Tholen taxonomy (Tholen, 1984) from the 8-color asteroid survey (Zellner et al., 1985), but recommended changes to Tholen's S-class representation when he incorporated the 52-color asteroid survey (Bell et al., 1988) into his supervised training model. This suggested for the first time that Artificial Neural Networks may be more a more sensitive classifier than Principal Component Analysis; however, most statistical asteroid spectral analyses are still

performed with Principal Component Analysis more than two decades later. The application of machine learning in this paper stems from the work in Howell et al. (1994)'s original approach and Mommert et al. (2016)'s more modern approach to machine learning.

Mommert et al. (2016) used various machine learning algorithms in an attempt to rapidly characterize new asteroid observations being taken with the Wide Field Camera for the United Kingdom Infrared Telescope (UKIRT) (WFCAM, Casali et al. 2007), across four bandpasses (Z, J, H, K) in the NIR. Mommert et al. (2016) utilized the Python module, scikit-learn (Pedregosa et al. 2011), which is the same module I include in my analysis in Section 4. Figure 3-3 is a flowchart of the various machine learning algorithms for a given dataset. There is a red box around the k-Nearest Neighbor algorithm, as it is the algorithm I implemented in this analysis. Classification machine learning algorithms take advantage of the binary nature of the output and employ supervised learning through the use of known training dataset outputs. The other commonly used method of assigning asteroid classifications, Principal Component Analysis, is located within dimensionality reduction subdivision of the machine learning flowchart. Principal Component Analysis reduces the number of dimensions through linear feature projection prior to running the machine learning algorithm to avoid the curse of dimensionality. Figure 3-3 only recommends using Principal Component Analysis when you are trying to explore the data, not make predictions.

The k-Nearest Neighbor algorithm learns parameter space from the training dataset and then applies a minimum distance calculation from the test point to the training dataset to determine classification. Figure 3-4 shows an example of the k-Nearest Neighbor method using radial distance. The k-value is the number of neighbors the method will look at when determining which classification group to put the test data point in. A higher k-value helps to eliminate outliers, but can be detrimental to the algorithm's ability to classify when using

Figure 3-3: Flowchart describing how to select an algorithm to solve a specific machine learning problem (https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html).
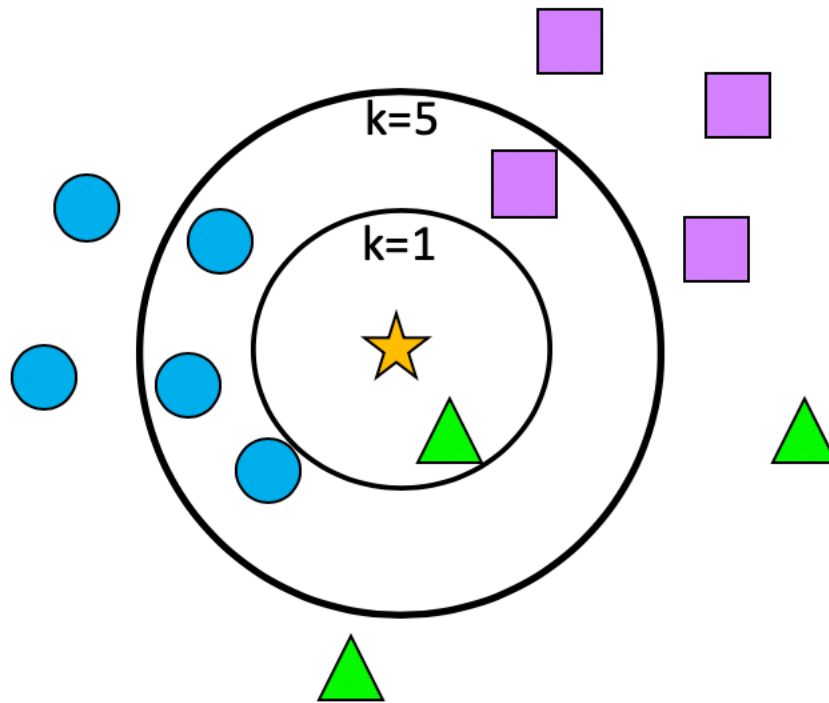
Figure 3-4: Example of k-Nearest Neighbor classification method with regions of k=1 and k=5 shown. The training dataset consists of the blue circles, purple squares, and green triangles, while the yellow star represents the test data.

using smaller training sets. In Figure 3-4, the yellow star represents the test data point to be classified and the green triangles, blue circles, and purple square represent the training dataset already "learned". If k = 1, the test data point would be classified as a green triangle, as a green triangle is the singular nearest training dataset point. If the classifier extended to encompass five neighbors, the algorithm would choose the blue circle class, as it has three neighbors in the distance region compared to one from the green triangle class and one from the purple square class.

A classification algorithm can be considered self-consistent if it can correctly predict which class the test data points should belong to based only on the information from the training dataset. It is imperative that these test data points have known outcomes, but not be

included in the training set. A technique for smaller datasets is sometimes to train the algorithm on all available data points and then remove one point to use as a test data point. This invalidates the self-consistent approach; independent test datasets should always be utilized. In the case of small sample sizes, a system of weights can be used and tests for the k-value that maximize accuracy across various test set sample sizes should be implemented.

4   OBSERVATIONS AND DATA ANALYSIS

The first step toward applying a machine learning technique to asteroid spectrophotometry is to build the training and test datasets that will teach the model. Larger training and test datasets typically result in more robust and accurate models. To that end, I obtained spectroscopic observations consisting of 277 observations of 241 asteroids from the Planetary Data System[2], to form an initial dataset for the analysis.

4.1 Observations

4.1.1 Near-Infrared Spectra

Five NIR datasets were selected from the Planetary Data System, all of which utilized the low- to medium-resolution SpeX (Rayner et al. 2003) cross-dispersion spectrograph and imager, part of NASA's 3-meter Infrared Telescope Facility (IRTF) on Mauna Kea, Hawaii. Datasets using SpeX were chosen for their $0.8 - 2.5$ μm wavelength coverage and ability to detect spectral features associated with surface composition and mineralogy.

---

[2] https://pds.nasa.gov/

The first dataset (Hardersen 2016), observed from 2001-2015, consists of NIR reflectance spectra for 68 Main Belt asteroids. The second dataset (Fieber-Beyer 2015), observed from 2006-2015, consists of NIR reflectance spectra for 52 Main Belt asteroids. Reddy was the principal observer for the remaining three datasets, spanning observations from 2001-2012. Reddy & Sanchez (2016) consists of NIR reflectance spectra for 90 Main Belt asteroids, Reddy (2010) for 40 near-Earth asteroids, and Reddy & Sanchez (2017) for 27 near-Earth and Mars-crossing asteroids.

4.1.2 Visible Spectra

Most of the spectral features associated with composition are in the NIR part of the spectrum, but, as described in Section 2, the visible wavelength range is also crucial to classification. I utilized existing visible-wavelength datasets, where available, for the 241 asteroids in our dataset and removed those without complementary visible and NIR observations from the analysis.

The Planetary Data System contained corresponding visible spectra for 208 of our initial observations from a combination of the 24-Color Asteroid Survey (Chapman, Gaffey, & McFadden 1993; Chapman & Gaffey 1979; McFadden et al. 1984), SMASS (Xu et al. 1995; Xu et al. 1996), the second phase of the Small Main-Belt Asteroid Spectrographic Survey (SMASSII) (Bus & Binzel 2002a; Bus & Binzel 2003), and the Small Solar System Objects Spectroscopic Survey (S3OS2) (Lazzaro et al. 2004; Lazzaro et al. 2006).

The visible spectra, at minimum, contain data in the range from $0.49 - 0.92$ μm, allowing the corresponding visible and NIR observations to be merged over the common 0.8 – 0.9 μm region and normalized to 0.55 μm. Figure 4-1 gives an example of the spectra for this dataset.
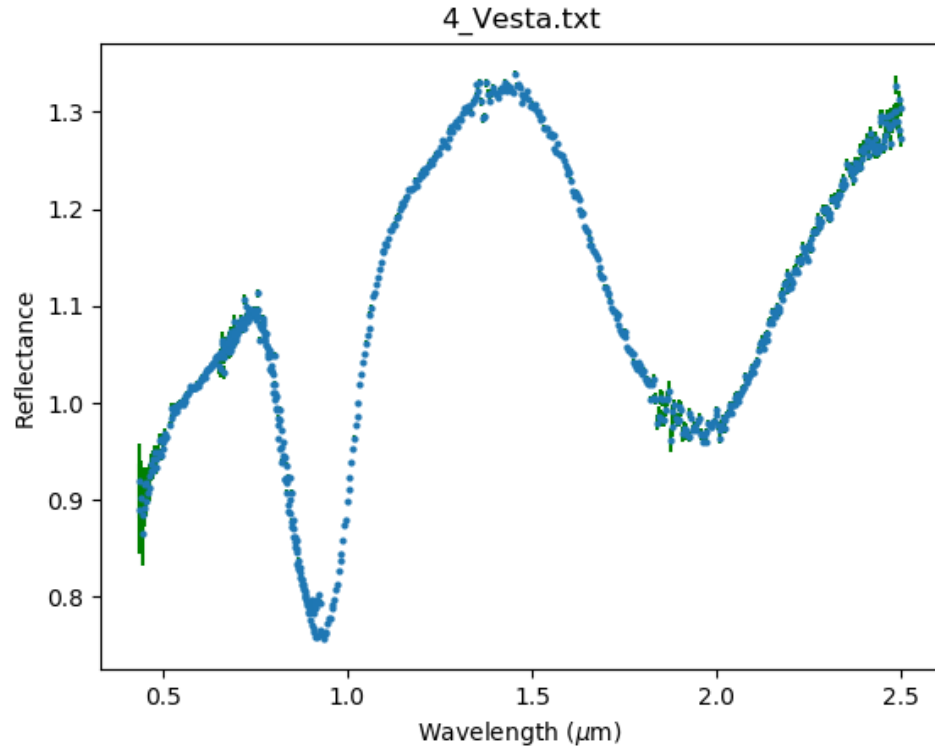
Figure 4-1: The concatenated visible and NIR spectra of (4) *Vesta*. The blue represents the data and the green is the 1-σ error region. The spectrum is normalized at 0.55 μm.

4.1.3 Assigning Bus-DeMeo Classes: χ^2 Thresholds

I used the online Bus-DeMeo taxonomy tool[3] to assign spectral classes to each of the remaining 208 observations in my dataset. To ensure a reasonable goodness of fit for each spectrum, I implemented a derivation of the χ^2 test (Bevington & Robinson 1992) to minimize the standard deviation of the errors:

$$\beta = \frac{1}{N} \sqrt{\Sigma_i^N ((x_i - \mu_i) - (\overline{x - \mu}))^2} \qquad (4.1)$$

---

where β is the quantity to minimize (the standard deviation of the errors), N is the number of

variables, x_i and μ_i are the values of the asteroid and mean taxonomic class, respectively, at

each wavelength, i, and $\overline{x - \mu}$ is the mean of the set of errors, $x_i - \mu_i$.

The data are well described when β → 0, so I set a cutoff threshold of β ≤ 0.01, using

a polynomial curve to smooth the spectra where necessary. Twenty-five observations, across

the S-, X-, and C-complexes, with β values over the threshold were removed from the final

dataset. See Table 4-1 for information on the 183 members of the final dataset.

Asteroid	Wavelength (µm)	β	Bus-DeMeo Classification
(1) Ceres	0.43-2.5[4,8]	0.0033	Ch
(2) Pallas	0.43-2.5[4,8]	0.0002	B
(4) Vesta	0.43-2.50[4,8]	0.0056	V
(9) Metis	0.33-2.50[1,8]	0.0025	S
(12) Victoria	0.43-2.5[4,8]	0.0022	S
(16) Psyche	0.43-2.5[4,6]	0.0036	X
(20) Massalia	0.43-2.5[4,8]	0.0016	S
(21) Lutetia	0.43-2.5[4,6]	0.0001	Xe
(22) Kalliope	0.43-2.5[4,8] 0.43-2.54[4,6]	0.0014 0.0029	X X
(30) Urania	0.43-2.50[4,8]	0.0005	S
(37) Fides	0.43-2.50[4,8]	0.0002	S
(41) Daphne	0.43-2.50[4,8]	0.0004	Ch
(44) Nysa	0.43-2.50[4,8]	0.0005	T
(45) Eugenia	0.43-2.50[4,8]	0.0028	C
(46) Hestia	0.43-2.50[4,5]	0.0085	Cb
(51) Nemausa	0.43-2.48[4,8]	0.0056	Cgh
(55) Pandora	0.43-2.50[4,6]	0.0004	T
(56) Melete	0.43-2.50[4,8]	0.0004	Cg
(63) Ausonia	0.43-2.50[4,8]	0.0087	Sw
(64) Angelina	0.43-2.50[4,8]	0.0002	T
(66) Maja	0.43-2.50[4,8]	0.0003	Ch
(69) Hesperia	0.43-2.50[4,6] 0.43-2.50[4,6]	0.0006 0.0013	Xc Xk

(71) Niobe	0.43-2.50[4,6]	0.0007	Xe
(77) Frigga	0.43-2.50[4,6]	0.0010	X
(84) Klio	0.43-2.50[4,8]	0.0002	Ch
(87) Sylvia	0.43-2.50[4,8]	0.0010	X
(97) Klotho	0.33-2.44[1,6]	0.0046	Xc
	0.33-2.47[1,6]	0.0049	Xc
(101) Helena	0.43-2.5[4,8]	0.0003	S
(105) Artemis	0.43-2.5[4,8]	0.0002	Cgh
(110) Lydia	0.43-2.5[4,6]	0.0040	Xk
	0.43-2.4[4,6]	0.0070	Xk
	0.43-2.5[4,6]	0.0095	Xk
(113) Amalthea	0.43-2.5[4,8]	0.0059	S
(121) Hermione	0.43-2.5[4,8]	0.0003	Ch
(125) Liberatrix	0.43-2.5[4,6]	0.0003	X
(129) Antigone	0.43-2.5[4,6]	0.0033	X
	0.43-2.5[4,6]	0.0005	X
(130) Elektra	0.43-2.5[4,8]	0.0003	Cgh
(132) Aethra	0.43-2.5[4,6]	0.0087	A
	0.43-2.5[4,6]	0.0082	A
(135) Hertha	0.43-2.5[4,8]	0.0005	Xk
	0.43-2.5[4,6]	0.0006	Xk
(136) Austria	0.43-2.5[4,6]	0.0022	Xe
(138) Tolosa	0.48-2.5[3,8]	0.0008	S
(167) Urda	0.43-2.5[4,8]	0.0010	S
(170) Maria	0.43-2.5[4,8]	0.0015	S
(182) Elsa	0.43-2.43[4,8]	0.0013	S
(184) Dejopeja	0.43-2.5[4,8]	0.0003	X
	0.43-2.48[4,6]	0.0007	X
(192) Nausikaa	0.43-2.5[4,8]	0.0036	Sq
(198) Ampella	0.43-2.5[4,5]	0.0035	Sv
(201) Penelope	0.43-2.5[4,6]	0.0015	Xc
	0.43-2.5[4,6]	0.0003	Xc
	0.43-2.5[4,6]	0.0006	X
(213) Lilaea	0.43-2.5[4,8]	0.0014	C
(214) Aschera	0.43-2.5[4,8]	0.0071	T
(216) Kleopatra	0.43-2.5[4,6]	0.0033	Xe
	0.43-2.5[4,6]	0.0078	Xe
(224) Oceana	0.49-2.46[2,6]	0.0084	X
(233) Asterope	0.43-2.5[4,8]	0.0003	Xk
(243) Ida	0.43-2.5[4,8]	0.0006	S
(248) Lameia	0.47-2.50[3,5]	0.0018	Xc
(250) Bettina	0.43-2.47[4,6]	0.0011	X

(253) Mathilde	0.43-2.4[4,8]	0.0034	C
(255) Oppavia	0.49-2.50[2,8]	0.0024	X
(256) Walpurga	0.49-2.50[3,8]	0.0004	X
(264) Libussa	0.43-2.50[4,8]	0.0027	S
(273) Atropos	0.49-2.50[2,8]	0.0008	X
(283) Emma	0.49-2.50[2,8]	0.0003	C
(289) Nenetta	0.43-2.50[4,8]	0.0078	A
(292) Ludovica	0.45-2.50[3,5]	0.0053	S
(306) Unitas	0.43-2.50[4,8]	0.0027	S
(308) Polyxo	0.43-2.50[4,8]	0.0006	T
(317) Roxane	0.43-2.50[4,8]	0.0005	T
(323) Brucia	0.49-2.50[2,9]	0.0013	S
(325) Heidelberga	0.33-2.50[1,6]	0.0021	X
(329) Svea	0.49-2.50[2,5]	0.0002	Cb
(335) Roberta	0.43-2.50[4,5]	0.0045	B
(338) Budrosa	0.43-2.50[4,6]	0.0003	Xk
(347) Pariana	0.33-2.44[1,6]	0.0008	X
(349) Dembowska	0.43-2.50[4,8]	0.0053	S
(355) Gabriella	0.43-2.50[4,5]	0.0076	S
(379) Huenna	0.43-2.50[4,8]	0.0027	C
(389) Industria	0.43-2.50[4,8]	0.0011	S
(391) Ingeborg	0.43-2.50[4,9]	0.0027	S
(403) Cyane	0.43-2.50[4,8]	0.0006	S
(413) Edburga	0.43-2.48[4,6]	0.0029	X
(417) Suevia	0.43-2.5[4,6]	0.0082	X
	0.43-2.49[4,6]	0.0045	X
(418) Alemannia	0.49-2.50[2,6]	0.0003	X
	0.49-2.50[2,6]	0.0004	X
(419) Aurelia	0.49-2.50[2,8]	0.0001	C
(434) Hungaria	0.43-2.50[4,8]	0.0002	T
(441) Bathilde	0.43-2.50[4,6]	0.0042	Xe
(442) Eichsfeldia	0.43-2.50[4,8]	0.0074	Ch
(446) Aeternitas	0.43-2.50[4,8]	0.0014	A
(458) Hercynia	0.43-2.50[4,8]	0.0005	S
(470) Kilia	0.43-2.50[4,8]	0.0023	S
(472) Roma	0.49-2.50[2,8]	0.0009	S
(495) Eulalia	0.45-2.50[3,5]	0.0018	B
(502) Sigune	0.49-2.50[2,8]	0.0094	S
(504) Cora	0.49-2.50[2,8]	0.0036	X
	0.49-2.47[2,6]	0.0015	X

	$0.49\text{-}2.50^{2,6}$	0.0005	X
(516) Amherstia	$0.43\text{-}2.50^{4,6}$	0.0005	X
(556) Phyllis	$0.43\text{-}2.50^{4,5}$	0.0023	S
(558) Carmen	$0.49\text{-}2.50^{2,6}$	0.0011	X
(569) Misa	$0.43\text{-}2.50^{4,8}$	0.0007	Ch
(619) Triberga	$0.49\text{-}2.50^{2,5}$	0.0040	Sv
(660) Crescentia	$0.49\text{-}2.50^{2,5}$	0.0093	S
(663) Gerlinde	$0.49\text{-}2.50^{2,8}$	0.0004	C
(670) Ottegebe	$0.43\text{-}2.50^{4,8}$	0.0010	S
(695) Bella	$0.33\text{-}2.50^{1,5}$	0.0046	S
(704) Interamnia	$0.43\text{-}2.50^{4,8}$	0.0010	B
(714) Ulula	$0.33\text{-}2.50^{1,5}$	0.0011	S
(739) Mandeville	$0.43\text{-}2.49^{4,6}$	0.0098	C
(741) Botolphia	$0.43\text{-}2.50^{4,8}$	0.0013	X
(758) Mancunia	$0.33\text{-}2.53^{1,6}$	0.0099	X
(762) Pulcova	$0.49\text{-}2.50^{2,8}$	0.0008	Cg
(785) Zwetana	$0.43\text{-}2.48^{4,6}$	0.0006	Cb
(787) Moskva	$0.47\text{-}2.50^{3,5}$	0.0006	S
(796) Sarita	$0.43\text{-}2.50^{4,6}$	0.0003	X
(797) Montana	$0.43\text{-}2.50^{4,5}$	0.0004	Sv
(809) Lundia	$0.49\text{-}2.50^{2,8}$	0.0075	V
(857) Glasenappia	$0.49\text{-}2.50^{2,6}$	0.0093	S
	$0.49\text{-}2.50^{2,6}$	0.0019	S
(858) El Djezair	$0.33\text{-}2.50^{1,8}$	0.0018	S
(860) Ursina	$0.43\text{-}2.50^{4,6}$	0.0002	X
	$0.43\text{-}2.50^{4,6}$	0.0006	X
(863) Benkoela	$0.43\text{-}2.50^{4,8}$	0.0048	A
(872) Holda	$0.43\text{-}2.50^{4,8}$	0.0003	X
	$0.43\text{-}2.44^{4,6}$	0.0003	X
(879) Ricarda	$0.48\text{-}2.50^{3,5}$	0.0052	Sv
(897) Lysistrata	$0.43\text{-}2.50^{4,5}$	0.0032	S
(908) Buda	$0.43\text{-}2.50^{4,5}$	0.0007	D
(951) Gaspra	$0.43\text{-}2.50^{4,8}$	0.0059	S
(1018) Arnolda	$0.49\text{-}2.5^{2,5}$	0.0056	S
(1036) Ganymed	$0.33\text{-}2.5^{1,9}$	0.0021	S
	$0.33\text{-}2.5^{1,5}$	0.0017	S
(1086) Nata	$0.44\text{-}2.5^{4,8}$	0.0007	Ch
(1145) Robelmonte	$0.45\text{-}2.5^{3,8}$	0.0020	S
(1166) Sakuntala	$0.45\text{-}2.5^{3,5}$	0.0003	S
(1215) Boyer	$0.49\text{-}2.5^{2,5}$	0.0008	S
(1251) Hedera	$0.43\text{-}2.5^{4,8}$	0.0004	X
(1284) Latvia	$0.43\text{-}2.5^{4,8}$	0.0017	T

(1329) Eliane	0.43-2.5[4,8]	0.0075	Sq
(1358) Gaika	0.45-2.5[3,5]	0.0003	Cg
(1379) Lomonosowa	0.45-2.5[3,5]	0.0005	S
(1459) Magnya	0.49-2.5[2,6]	0.0051	V
(1501) Baade	0.43-2.5[3,5]	0.0005	Sv
(1587) Kahrstedt	0.43-2.5[4,5]	0.0019	Sr
(1607) Mavis	0.48-2.5[3,5]	0.0089	Sq
(1620) Geographos	0.33-2.5[1,7]	0.0010	S
(1626) Sadeya	0.49-2.5[3,8]	0.0012	Sv
(1717) Arlon	0.33-2.5[1,8]	0.0062	Sq
(1722) Goffin	0.49-2.5[3,5]	0.0022	S
(1727) Mette	0.33-2.5[1,9]	0.0038	S
(1772) Gagarin	0.45-2.5[3,5]	0.0031	S
(1830) Pogson	0.43-2.46[4,8]	0.0004	S
(1854) Skvortsov	0.49-2.5[3,5]	0.0019	S
(1883) Rimito	0.49-2.5[2,8]	0.0005	S
(1929) Kollaa	0.43-2.5[4,8]	0.0046	V
(1980) Tezcatilpoca	0.43-2.5[2,7]	0.0091	Sw
(2011) Veteraniya	0.49-2.5[3,8]	0.0013	V
	0.49-2.54[3,6]	0.0020	V
(2014) Vasilevskis	0.47-2.5[3,8]	0.0018	S
(2035) Stearns	0.43-2.5[4,9]	0.0006	Xk
(2045) Peking	0.43-2.5[4,8]	0.0009	V
(2089) Cetacea	0.43-2.5[4,5]	0.0020	Sq
(3066) McFadden	0.49-2.50[2,5]	0.0004	S
(3345) Tarkovskij	0.43-2.50[4,5]	0.0059	C
(3760) Poutanen	0.49-2.50[3,5]	0.0024	S
(3782) Celle	0.43-2.54[4,6]	0.0059	V
	0.43-2.53[4,6]	0.0015	V
(3849) Incidentia	0.43-2.54[4,6]	0.0019	V
	0.43-2.54[4,6]	0.0024	V
(3999) Aristarchus	0.49-2.50[3,5]	0.0061	C
(4179) Toutatis	0.49-2.45[3,7]	0.0032	Sq
	0.49-2.50[3,9]	0.0001	S
(4954) Eric	0.49-2.54[2,7]	0.0018	S

Table 4-1: List of the asteroids, observations, β values, and initial Bus-DeMeo classifications used in this analysis. [1]Chapman, Gaffey, & McFadden 1993; Chapman & Gaffey 1979; McFadden et al. 1984. [2]Lazzaro et al. 2004; Lazzaro et al. 2006. [3]Xu et al. 1995; Xu et al. 1996. [4]Bus & Binzel 2002a; Bus & Binzel 2003. [5]Fieber-Beyer 2015. [6]Hardersen 2016. [7]Reddy 2010. [8]Reddy & Sanchez 2016. [9]Reddy & Sanchez 2017.

4.2 Spectrophotometry

Asteroid photometry, while lower resolution than spectroscopy, can still provide clues about the physical properties and surface composition of asteroids despite not being able to distinguish individual spectral features. This section details the methodology applied to derive spectrophotometric values from spectroscopic observations.

4.2.1 Photometric Band Selection

As described in Section 2.2, the surface properties of asteroids alter their resulting reflectance spectra. Photometric filters can be strategically utilized to identify distinct features of asteroid classes in the visible and NIR regions. Figure 4-2 and Figure 4-3 show normalized
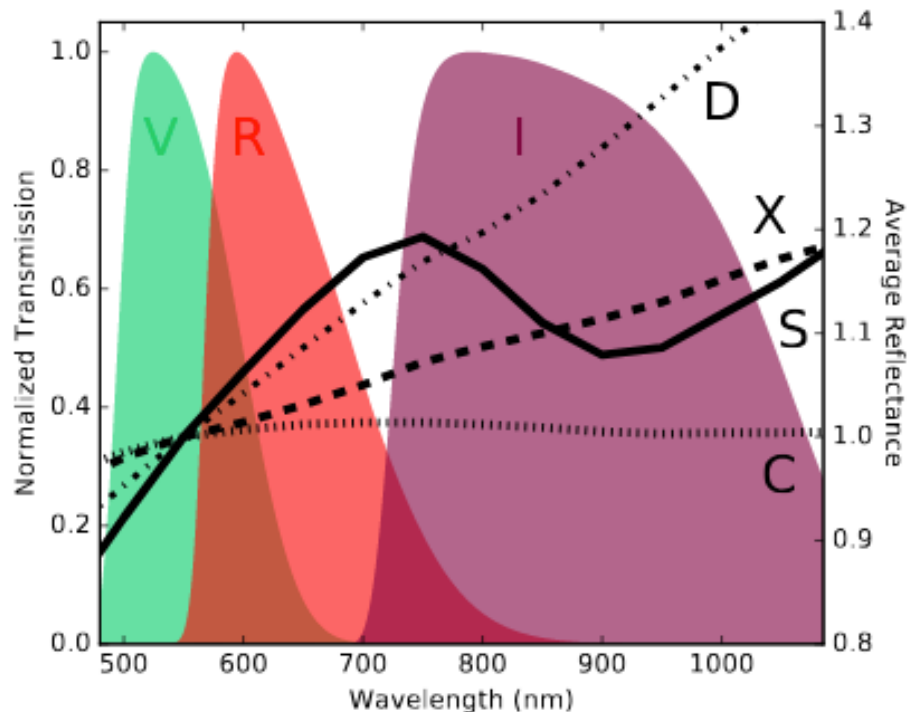
Figure 4-2: V, R, and I Johnson-Cousins filters overplotted with the averaged asteroid reflectance spectra from the Bus-DeMeo taxonomy, normalized at 0.55 μm. Extracted from Figure 2 of Erasmus et al. 2018.
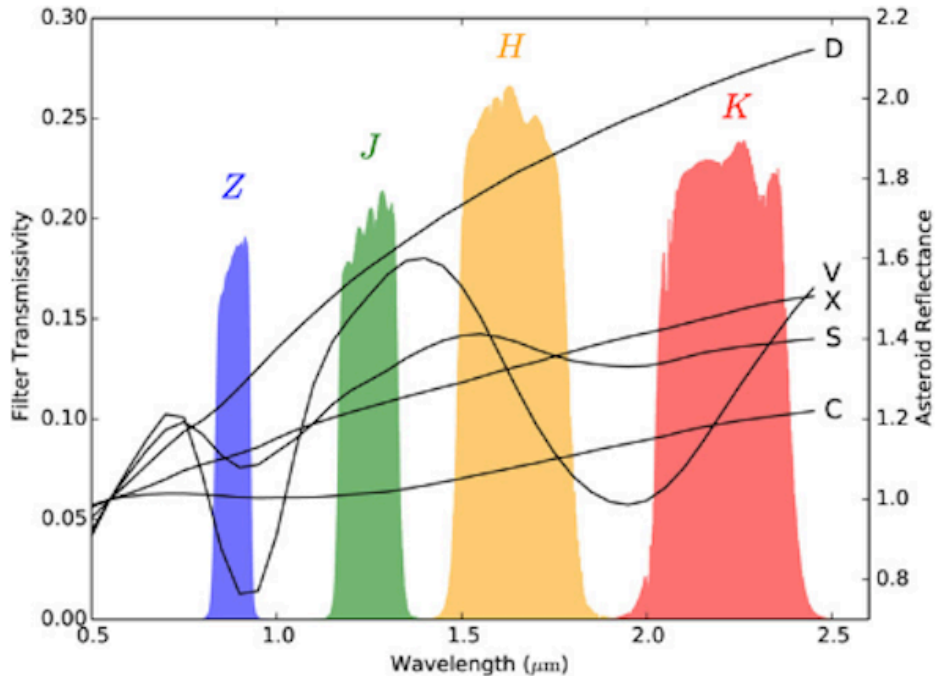
Figure 4-3: Z, J, H, and K bandpasses overplotted with the averaged asteroid reflectance spectra from the Bus-DeMeo taxonomy, normalized at 0.55 μm. Extracted from Figure 1 of Mommert et al. 2016.

transmission of common filter bandpasses across the visible and NIR wavelength regions, respectively. The three main complexes (S, X, and C) from the Bus-DeMeo taxonomy, as well as the D and, for the NIR, V endmembers groups have been overplotted with their mean spectral reflectance values.

In Figure 4-2, the differences between the taxonomic classes can be easily seen from one bandpass to another. There is an increase in slope for all types except the C-complex, making that a diagnostic choice for the C-complex. The D-type has a drastic increase in slope, while overall slope may not prove very diagnostic for the X- and S-complexes. To diagnose the X- and S-complexes, the more short-term changes due to the spectral features in the S-complex that are not present in the X-complex can be noted. A major weakness in relying on the overall slope of a spectrum in determining taxonomic class is that it can be affected by space weathering and particle size.

In Figure 4-3, the same methodology can be applied to the NIR region. In this region, the Z-band can diagnose the 1 μm olivine and pyroxene features, while the K-band can diagnose the 2 μm pyroxene feature. The J-band can help to diagnose the overall slope as, unlike the other three bands, the slope of all the taxonomic classes shown are relatively constant.

4.2.2 Vega Magnitude Calculation

A value for the photometric flux must be derived to calculate the color magnitudes of the asteroids across the desired bandpasses (Mommert et al. 2016; Equation 1):

$$F_\lambda = \int T_{f,\lambda} A_\lambda S_\lambda \lambda \, d\lambda \qquad (4.2)$$

where F_λ is the photometric flux across the wavelength range, λ , $T_{f,\lambda}$ is the transmittance of the bandpass filter across λ, A_λ is the reflectance spectrum of the asteroid across λ, and S_λ is the spectrum of the Sun[4] across λ. Table 4-2 lists the bandpasses in this analysis, along with the effective wavelength, central wavelength, and standard filter used to calculate the flux. A spline interpolation was performed on the solar and asteroid spectra before convolving the filter throughput with the product of the two spectra across the wavelength range of the bandpass.

The resulting flux is utilized to calculate the Vega magnitude (Bessell & Murphy 2012):

$$V_{mag} = -2.5 \log(F_\lambda) + Zero\ Point \qquad (4.3)$$

where V_{mag} is the Vega magnitude value, F_λ is the photometric flux calculated in Equation 4.2, and the Zero Point is a correction factor based on the magnitude of the star, Vega[5], being zero

---

[4] http://kurucz/harvard.edu/stars/Sun/fsunallp.1000

[5] http://kurucz.harvard.edu/stars/vega/vegallpr25.1000

in all filters. To calculate the Zero Point, I used Equation 4.2, substituting the spectrum of Vega for S_λ and setting A_λ to unity. The resulting magnitudes are used to calculate the color indices in the following section.

Standard Filter	λ_{eff} (μm)	FWHM (μm)	Reference
V	0.551	0.088	Johnson (1966)
R	0.658	0.138	Cousins (1976)
I	0.806	0.149	Cousins (1976)
Z	0.905	0.137	Fukugita et al. (1996)
J	1.220	0.213	Bessell (2005)
K	2.190	0.390	Bessell (2005)

Table 4-2: Bandpass values used in this analysis.

4.2.3 Color Indices

An advantage to using color indices for classification in the visible and NIR is the ability to probe the information of both the 1 and 2 μm features and the overall slope of the spectrum. In the Bus-DeMeo taxonomy, the slope is removed prior to any analysis (DeMeo et al. 2009), despite having diagnostic information about particle size and space weathering (Bus et al. 2002). Selecting color indices allows meaningful associations across the spectrum to be made. In the visible region, I selected V-R and V-I as my two components. V-R probes short-term changes in the slope of the spectrum while V-I represents the change in slope across the visible wavelength region. In the NIR region, the two colors chosen are J-K and Z-J. J-K determines an overall change in slope as well as probing the 2 μm feature and Z-J probes for the 1 μm

feature and more abrupt changes in slope. The final correction made before determining color

values is to correct for the solar colors, listed in Table 4-3.

Color Index	Correction	Reference
V-R	0.352	Ramirez et al. (2012)
V-I	0.702	Ramirez et al. (2012)
J-K	0.362	Casagrande et al. (2012)
Z-J	0.369	Mommert et al. (2016)

Table 4-3: Solar color correction factors.

4.3 Implementing Machine Learning Method using Python

4.3.1 Training and Test Dataset Selection

There are two basic ways to setup a machine learning dataset: either for supervised or

unsupervised learning. Supervised learning requires a dataset with many features that the

algorithm explores in order to learn the dataset's properties, while unsupervised learning

utilizes a dataset with a labeled outcome. Classification methods, including the one described

here, use a supervised learning algorithm. To that end, I now have a dataset of 183

observations, each containing two sets of photometric colors associated with a known Bus-

DeMeo taxonomic class. This dataset is divided into a training dataset for the model and an

independent test dataset. A member of the test dataset is one whose outcome is known but was

not included in the training set (Howell et al. 1994). Given the relatively small sample size, 37

(20%) observations were used for the test set and the remaining 146 (80%) observations were

used for the training set. It is important for machine learning algorithms to learn on a large,

diverse dataset to improve the accuracy but a representative test dataset is crucial as well. Test

dataset sizes were explored from 1-90% and the balance between accuracy and a representative sample was determined to be 20% of the total dataset.

4.3.2 k-Nearest Neighbor Algorithm

I selected the k-Nearest Neighbor algorithm due to the classification nature of the problem, the discrete nature of the outcomes, and the relatively small dataset. The concept of the k-Nearest Neighbor algorithm and the scikit-learn Python package are described in Section 3.3.

The value of k was selected by sampling values of k = {1…9} and determining which value for k maximized the accuracy of the test dataset. This occurs when the accuracy of the testing dataset surpasses the accuracy of the training dataset. Typically, an odd value for k is selected to avoid the small possibility of a tie. Figure 4-4 shows the results of the k-varying test. The k-value that maximized accuracy for this dataset is k = 5.
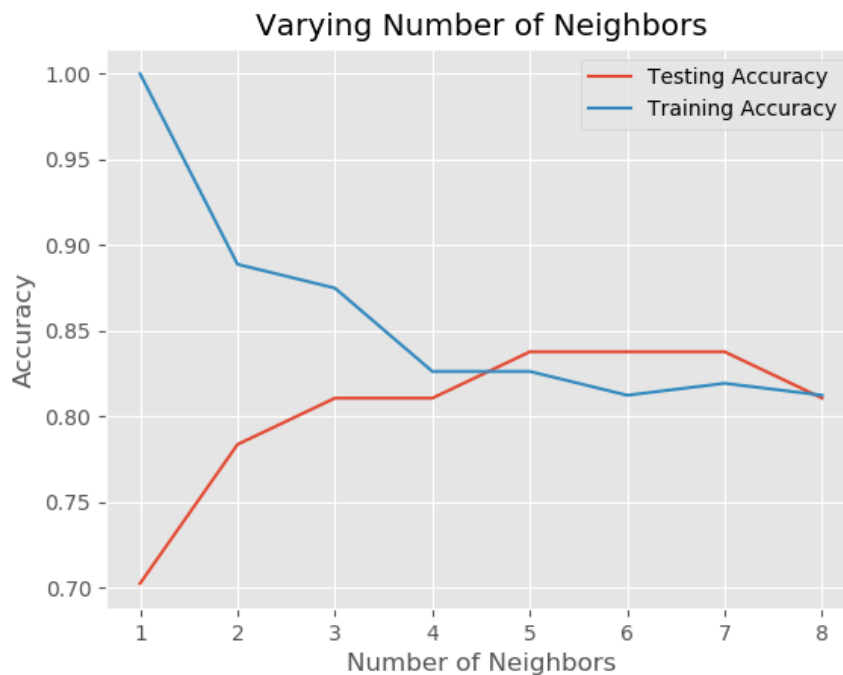

Figure 4-4: Results of a sampling test varying the number of neighbors, k, required to select a classification outcome. The testing accuracy surpasses the training accuracy at k=5.

A weighted distance metric to the closest point to the test point is used to establish classification. In this analysis, a Minkowski distance function was chosen as it tends to be more accurate for smaller, discrete datasets.

$$D\ (X,Y) =\ (\textstyle\sum_{i=1}^{n}|x_i -\ y_i|^p)^{1/p} \tag{4.4}$$

where D is the distance of between points x_i and y_i in sets X and Y with order p. In this analysis, I found the best accuracy was obtained with a value of p = 2, resulting in the Minkowski distance function behaving as a Euclidean distance function. Points are weighted according to the error on each color.

4.3.3 Interpreting Metrics

The simplest measure of the performance, and therefore error, of the machine learning algorithm is the accuracy of the model. Accuracy is the fraction of the test dataset for which the model produced the correct output. This is a very straightforward metric, but the power of measuring accuracy is that it encompasses all smaller, internal quantities. For example, when interpreting the k-Nearest Neighbor algorithm, should the penalty be higher for frequent medium-sized mistakes, such as being placed at the boundary of a neighboring classification, or for rare large mistakes, such as classifying one member into a completely unrelated outcome? Accuracy measurements adequately portray the effectiveness of a supervised, discrete, classification algorithm.

There are other metrics to determine the successfulness of the machine learning algorithm. Precision, recall, and the F_1 score are also used in statistical analyses when there is

binary classification. Precision refers to the proportion of the positive classifications that was

correct:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$ (4.5)

Recall is the metric of how many true positives were correctly identified:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$ (4.6)

A machine learning algorithm will return a high precision if there are considerably

more significant results than insignificant ones. Recall is directly related to the amount of

significant results being returned. The final metric utilized in this work is the F_1 score. The

F_1 score is a way to combine (and weight) the precision and the recall to compute an

accuracy score based on the harmonic mean of the two:

$$F_1 = (1 + \Phi^2) \times \frac{Precision \times Recall}{(\Phi^2 \times Precision) + Recall}$$ (4.7)

where the Φ-parameter determines the weight of precision in the final F_1 score. In this

analysis, the Φ-parameter is unity.

5 RESULTS AND DISCUSSION

This chapter contains a discussion of the results of the machine learning classification algorithm. The final dataset contained 27 C-complex asteroids, 70 S-complex asteroids, 57 X-complex asteroids, 11 V-type asteroids, 5 A-type asteroids, 8 T-type asteroids, 4 B-type asteroids, and 1 D-type asteroid. When asteroids were assigned to a sub-class of a complex: Cb, Cg, Cgh, Ch for C-complex, Sa, Sq, Sqw, Sr, Srw, Sv, Sw for S-complex, and Xc, Xe, Xk for X-complex, they were simplified into the larger complex. This could lead to a decrease in accuracy but was necessary given the size of the dataset. Note: the single asteroid class (D-type) was not included in the test dataset, but was included in the training dataset for use in future studies.

5.1 Visible Dataset

Figure 5-1 shows the training dataset for the visible region with the decision boundaries determined by the k-Nearest Neighbor algorithm. There are six distinct boundary regions, although the boundary regions are not completely smooth. Most notably, the T-type boundary region is located within the X-type, with other T-type members on the boundary between the X- and C-types and X- and S-types. This is indicative that either the T-type asteroids are easily misidentified by the Bus-DeMeo taxonomy, or more likely, that the relatively featureless spectra are not diagnostic in the visible wavelength region.
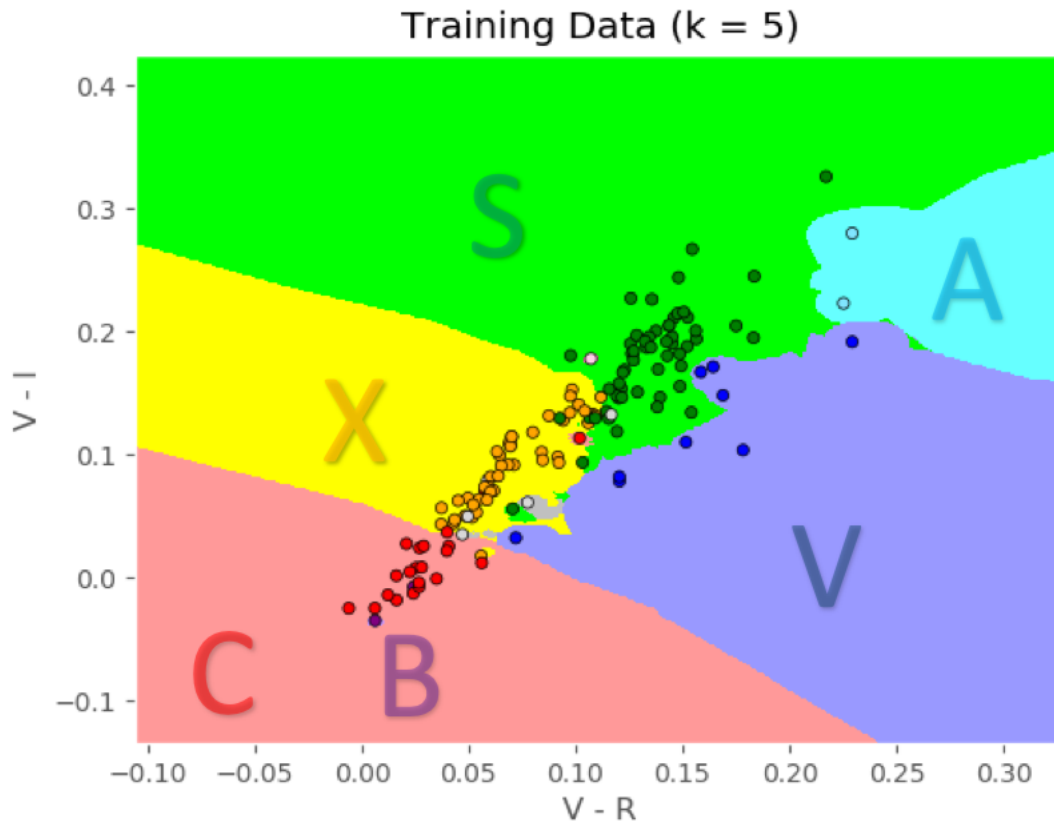
Figure 5-1: The training dataset using the V-I and V-R features plotted with the resulting decision boundaries created by the k-Nearest Neighbors algorithm. For this and following color maps, red represents the C-class asteroids, green represents the S-class asteroids, dark blue represents the V-type asteroids, light blue represents the A-type asteroids, yellow represents the X-class asteroids, silver represents the T-type asteroids, purple represents the B-type asteroids, and brown represents the D-type asteroids.

There is also a noticeable intrusion of the V-type asteroids into the area of the S-complex. This could be due to a misclassification of V-types, or a potential spectral alteration effect due to space weathering of surface minerals reddening the slopes.

To determine if our machine "learned" successfully, I calculated the metrics described in Section 4.3.3; the results are listed in Table 5-1. The overall accuracy of the test dataset is 83%. Accuracy scores over 80% are viewed as a successful proof-of-concept, while scores

over 95% are viewed as successful for classification. While a score of 83% implies the method is valid, it expects 17% (or ~6 members of our test dataset) to be misclassified.

Metric	Average Score
Precision	0.82
Recall	0.84
F_1 score	0.82
Overall Accuracy	0.83

Table 5-1: Classification report for the visible dataset

Figure 5-2 shows the results of the test dataset superimposed on the decision boundaries determined during the training dataset. Visually, the test dataset agrees with the decision boundaries with ~6 misclassifications near their correct boundaries. The Bus-DeMeo taxonomy was performed over the visible and NIR wavelength ranges, but the exclusion of slope may be leading to misclassifications in the visible portion of the spectrum.

There are several small end member class (T, D, B) boundaries within the boundary regions of the larger complexes. This led to the investigation of the S-, C-, and X-classes without the smaller end member classes, as seen in Figures 5-3 and 5-4 and Table 5-2. The overall precision increases to 94% when the end members are removed, moving the results from just above a proof-of-concept to 1% short of an accurate classification tool. This highlights the importance of number of free parameters versus sample size in machine learning. A reduction in more than half of the number of parameters with a moderate decrease in the sample size led to a substantial increase in accuracy. However, the exclusion of the end members can lead to generalizations and misclassifications of future datasets and such cases should be noted appropriately.

Figure 5-2: The test dataset using the V-I and V-R features plotted with the decision boundaries created by the k-Nearest Neighbors algorithm during the training dataset overlaid.
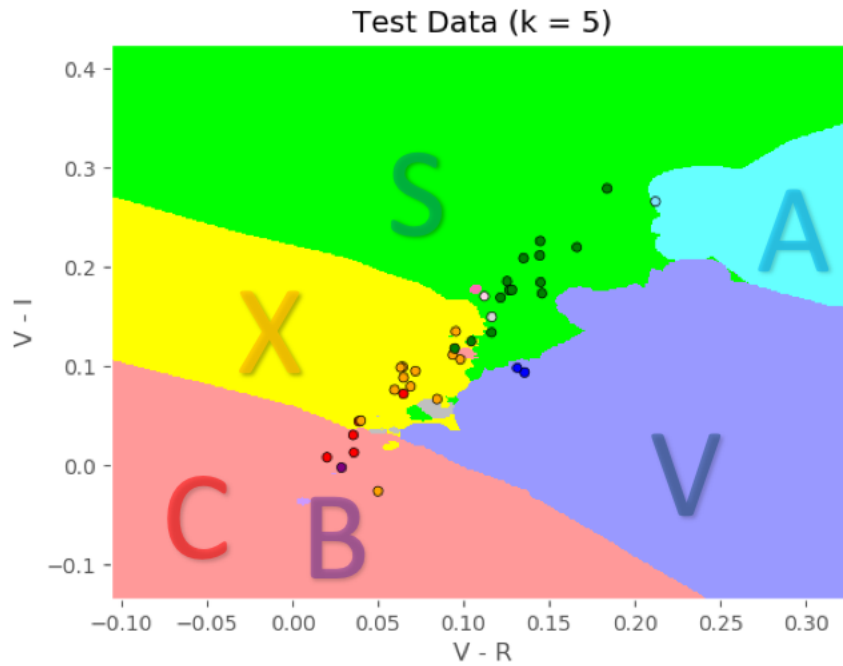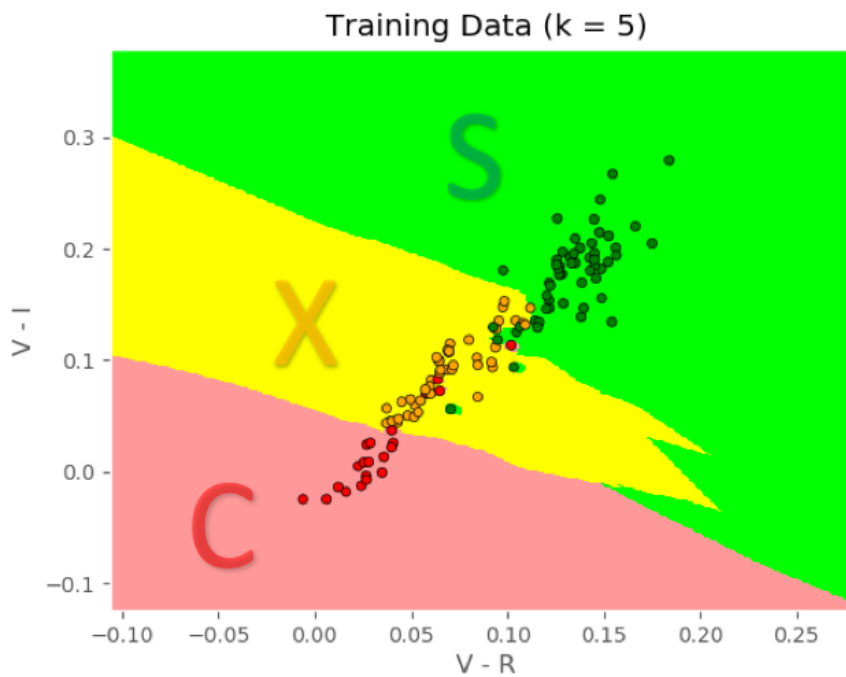

Figure 5-3: The training dataset using the V-I and V-R features plotted with the resulting decision boundaries created by the k-Nearest Neighbors algorithm. Red represents the C-class asteroids, green represents the S-class asteroids, and yellow represents the X-class asteroids.

Figure 5-4: The test dataset using the V-I and V-R features of the S-, C-, and X-classes plotted with the decision boundaries created by the k-Nearest Neighbors algorithm during the training dataset overlaid.

Metric	Average Score
Precision	0.94
Recall	0.94
F_1 score	0.94
Overall Accuracy	0.94

Table 5-2: Classification report for the visible dataset, using the S-, C-, and X-class asteroids.

5.2 NIR Dataset

Figure 5-5 shows the training dataset for the NIR region with the decision boundaries determined by the k-Nearest Neighbor algorithm. There are seven distinct boundary regions, which appear less smooth than in the visible region. The T-types do not have a distinct boundary region in the NIR, instead they are in the middle of the X-complex around (0.1, 0.05). This is reasonable as the T-type asteroids feature a sharp increase in slope in the visible portion of the spectrum, followed by a relative featureless, slightly reddened spectrum in the NIR. Therefore, the T-type is diagnostic in the visible portion of the spectrum but not in the NIR. The A-types, with their strong olivine feature, and V-types with their distinctive pyroxene feature, have very dominant boundary regions. The boundary separation of the A- and V-types combined with the complete lack of boundary region for the T-types validates the hypothesis that the NIR is more diagnostic of surface mineralogy features. The X-class has a very weak boundary region between the S- and C-classes, suggesting again that further investigation into the three class regions is necessary.

Results for the NIR dataset classification report are in Table 5-3. The overall accuracy of the test dataset is 78%, or ~8 members of our test dataset should be expected to be misclassified. The visible and NIR reports have similar accuracies, showing that the technique and dataset were maximized, and additional increases in accuracy will only come with more observations.

Figure 5-5: The training dataset using the Z-J and J-K features plotted with the resulting decision boundaries created by the k-Nearest Neighbors algorithm.

Figure 5-6: The test dataset using the Z-J and J-K features plotted with the decision boundaries created by the k-Nearest Neighbors algorithm during the training dataset overlaid.
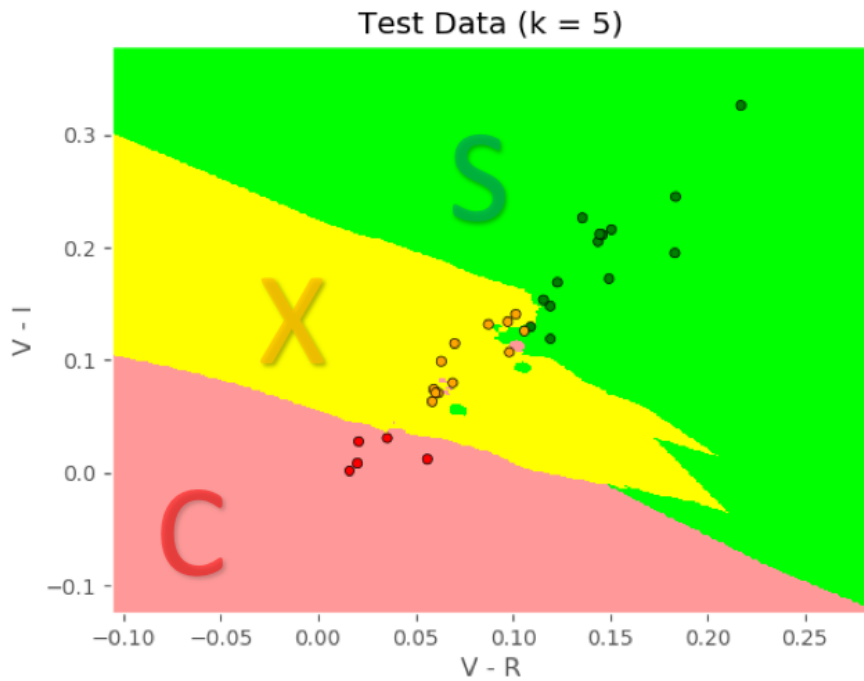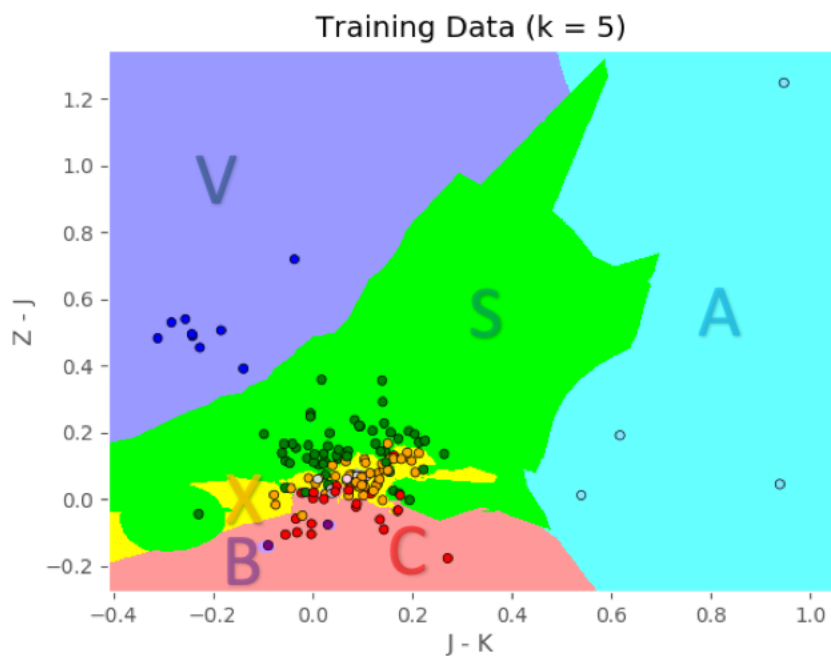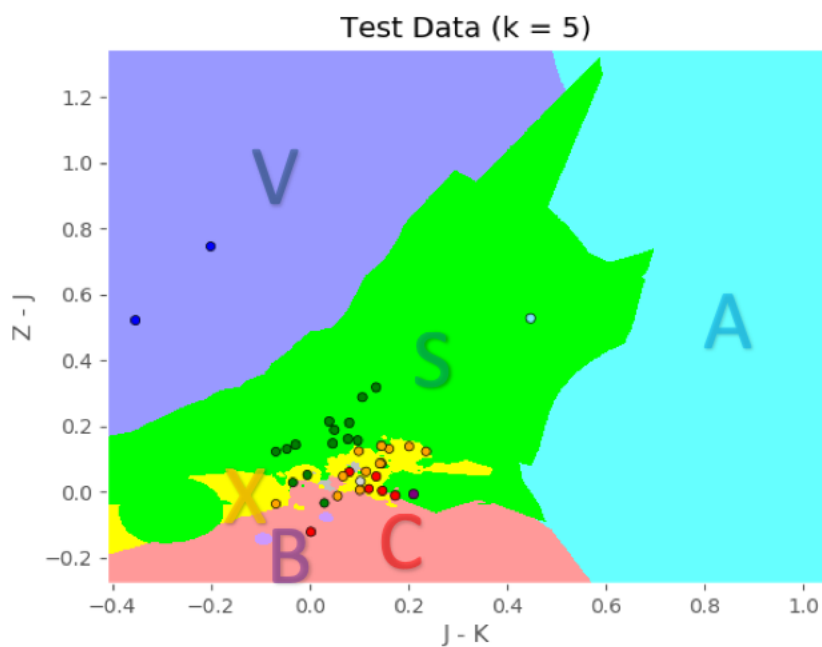
Metric	Average Score
Precision	0.77
Recall	0.80
F_1 score	0.77
Overall Accuracy	0.78

Table 5-3: Classification report for the NIR dataset

The test dataset clusters nicely into the boundary regions in Figure 5-6, except for a few data points in the central region. This central clustering of different taxonomic types was also seen in the training dataset, so it is not surprising to see in the test dataset. The one A-type asteroid in the test set is located within the S-class boundary region, but given the positions of the training A- and S-type asteroids, this may not be out of place. The Z-J color is probing the 1 μm olivine and pyroxene features, so the results of higher Z-J colors for S-complex and A-type with the highest values for the V-types is expected. The J-K color is probing the 2 μm pyroxene feature, and so the relatively vertical structuring of the C-, X-, and S-complexes follows this trend. The J-K color variation for the A-types may be due to the shallow 2 μm feature, or more likely the abrupt change in spectral slope from very steep to relatively flat.
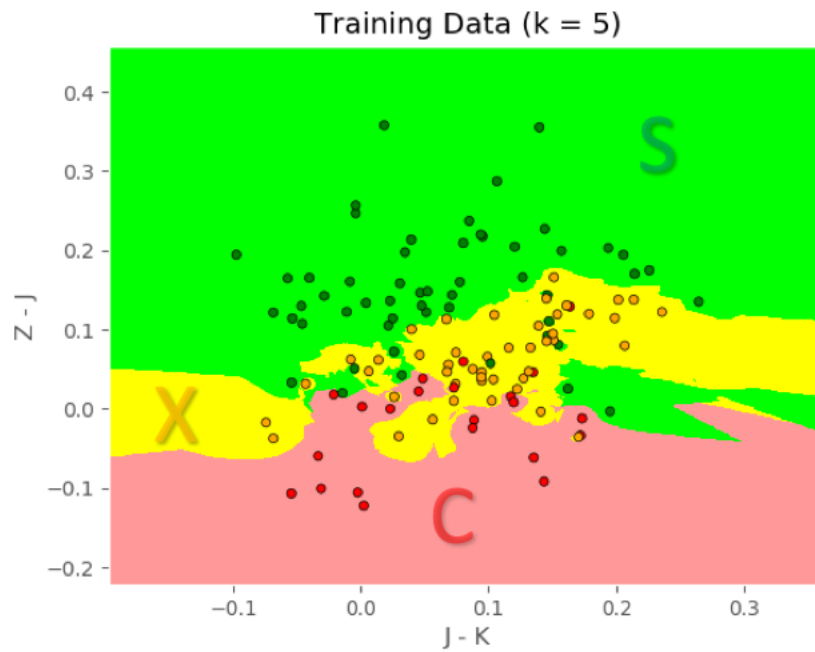
Figure 5-7: The training dataset using the Z-J and J-K features of the S-, C-, and X-classes plotted with the resulting decision boundaries created by the k-Nearest Neighbors algorithm.
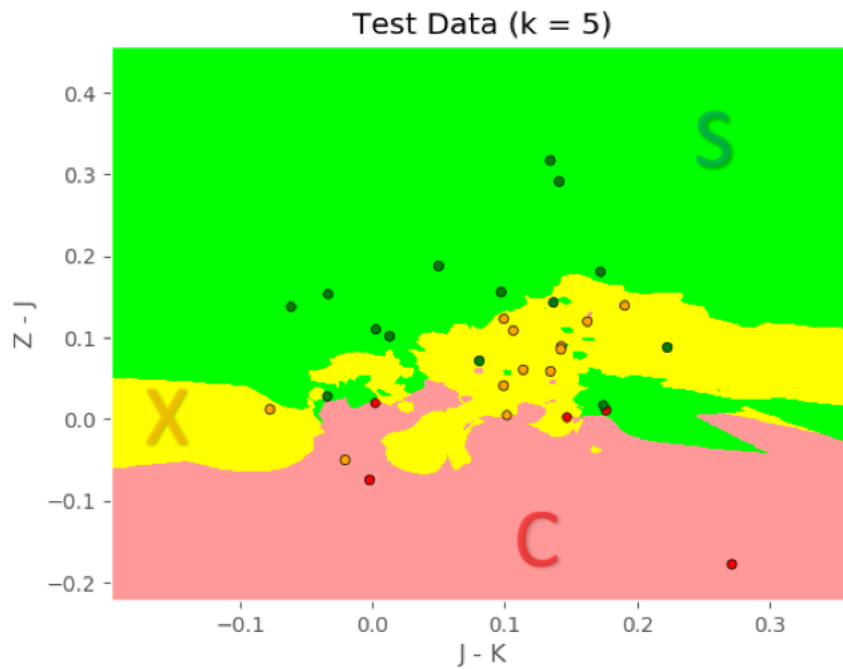
Figure 5-8: The test dataset using the Z-J and J-K features plotted with the decision boundaries created by the k-Nearest Neighbors algorithm during the training dataset overlaid.

Metric	Average Score
Precision	0.92
Recall	0.91
F_1 score	0.91
Overall Accuracy	0.91

Table 5-4: Classification report for the NIR dataset of the S-, C-, and X-classes.

Examining only the S-, C-, and X-classes improves the overall accuracy from 78% to 91%, but the boundary region of the X-class is still not well defined. Both the training dataset and test dataset for the S-, C-, and X-classes in the NIR (Figures 5-7 and 5-8) show clear boundaries between the S- and C-classes, but the X-class shows intrusions of both the S- and C- classes. The overall accuracy did improve to satisfy as a proof of concept, which the original NIR dataset did not, but it is further removed from use as an accurate classification schema than the visible dataset.

6   CONCLUSIONS AND FUTURE WORK

The primary goal of this analysis was to increase our understanding of asteroid properties through a combination of past perspectives and current analytical techniques. I achieved that goal through the implementation of the k-Nearest Neighbor machine learning algorithm on a dataset of spectrophotometric color indices derived from combined visible and NIR observations and paired with Bus-DeMeo taxonomic classes. The overall accuracy scores of the machine learning test dataset validates the methodology, while the results support that the visible wavelength region is more diagnostic of spectral slope and the NIR wavelength region is more diagnostic for surface mineralogy. However, both the complete datasets and reduced S-, C-, and X-complex datasets fall short of the accuracy necessary to replace the current method of taxonomic classification with machine learning. The general robustness of the Bus-DeMeo taxonomy is corroborated through the relatively similar grouping structure between the C-, S-, and X-complexes in both wavelength ranges, suggesting an overall relationship between slope and features present across multiple wavelength regions. This is possibly due to spectral features being closely tied to surface mineralogy and spectral slope alterations believed to be tied to the effects of space weathering.

This project was designed to be a test for the accuracy of the k-Nearest Neighbor machine learning algorithm when applied to spectrophotometric datasets. This technique is intended to be applied in the future to build a free, open-source online tool dedicated to utilizing machine learning for physical asteroid properties, such as size, albedo, or density as

well as applied for additional classification efforts. A focus should be placed on compiling

and training large datasets of spectroscopic and photometric observations for improving the

accuracy of these machine learning tools already at our disposal. By making training sets

available for the public, the computational cost of statistical analysis could be greatly

reduced, leading to new ideas and discoveries throughout the field.

REFERENCES

Adams, J. B., and McCord, T. B. 1971. Optical properties of mineral separates, glass, and anorthositic fragments from Apollo mare samples. *Proceedings Lunar Sci. Con.* **2**, 2183-2195.

Adams, J. B. 1974, Visible and near-infrared diffuse reflectance spectra of pyroxenes as applied to remote sensing of solid objects in the solar system, *J. Geophys. Res.* **79**, 4829-4836.

Barucci, M. A., Capria, M. T., Coradini, A., Fulchignoni, M. 1987. Classification of asteroids using G-mode analysis. *Icarus* **72**, 304-324.

Bell, J.F., Owensby, P.D., Hawke, B.R., & Gaffey, M.J. 1988. The 52-Color Asteroid Survey: Final Results and Interpretation. *Lunar and Planetary Inst. Technical Report* **19**, 57.

Bell, J.F., Davis, D.R., Hartmann, W.K., and Gaffey, M.J. 1989. Asteroids: The big picture. In: Binzel, R. P., Gehrels, T., Matthews, M. S. (eds.), *Asteroids II*. Univ. of Arizona Press, Tuscon, 921-948.

Bessell, M.S. 2005. Standard photometric systems. *ARA&A* **43**, 293-336.

Bessell, M. and Murphy, S. 2012. Spectrophotometric libraries, revised photonic passbands, and zero points for UBVRI, Hipparcos, and Tycho photometry. *PASP* **124**, 140-157.

Bevington, P.R. and Robinson, D. K. 1992. *Data Reduction and Error Analysis for the Physical Sciences*. New York: McGraw-Hill, Inc. 3rd ed.

Birlan, M., Barucci, M. A., and Fulchignoni, M. 1996. G-mode analysis of the reflection spectra of 84 asteroids. *A&A* **305**, 984-988.

Boss, A. P. 1997. Giant planet formation by gravitational instability. *Science* **276**, 1836-1839.

Boss, A. P. 1998. Evolution of the solar nebula. IV. Giant gaseous protoplanet formation. *ApJ* **503**, 923-937.

Bowell, E., Chapman, C. R., Gradie, J. C., Morrison, D., and Zellner, B. 1978. Taxonomy of asteroids. *Icarus* **35**, 313-335.

Brunetto, R. and Strazzulla, G. 2005. Elastic collisions in ion irradiation experiments: A mechanism for space weathering of silicates. *Icarus* **179**, 265–273.

Brunetto R., Romano F., Blanco A., Fonti, S., Martino, M., Orofino, V., and Verrienti, C. 2006. Space weathering of silicates simulated by nanosecond pulse UV excimer laser. *Icarus* **180**, 546–554.

Brunetto R., Loeffler M.J., Nesvorny D., Sasaki S., and Strazzulla G. 2015. Asteroid surface alteration by space weathering processes. In P. Michel et al. (eds.), *Asteroids IV*. Univ. of Arizona Press, Tucson, 597-616.

Burbine, T. H., Buchanan, P. C., Dolkar, T., and Binzel, R.P. 2009. Pyroxene mineralogies of near-Earth vestoids. *Meteorit. Planet. Sci.* **44**, 1331-1341.

Burbine, T.H. 2016. Advances in determining asteroid chemistries and mineralogies. *Geochemistry* **76**, 181-195.

Burns, R. G. 1970. *Mineralogical Applications of Crystal Field Theory*. London and New York: Cambridge Univ. Press. 224.

Burns R. C., Huggins F. E., Abu-Eid R. M. 1972. Polarized absorption spectra of single crystals of lunar pyroxene and olivine. *Moon* **4**, 93-102.

Bus, S.J. and Binzel, R.P. 2002a. Phase II of the small main-belt asteroid spectroscopic survey: The observations. *Icarus* **158**, 106-145.

Bus, S.J. and Binzel, R.P. 2002b. Phase II of the small main-belt asteroid spectroscopic survey: A feature-based taxonomy. *Icarus* **158**, 146-177.

Bus, S.J., Vilas, F. and Barucci, M.A. 2002. Visible-wavelength spectroscopy of asteroids. In Bottke et al. (eds.), *Asteroids III*, Univ. of Arizona Press, Tucson. 169-182.

Casagrande, L., Ramirez, I., Melendez, J., and Asplund, M. 2012. The infrared colors of the sun. *ApJ* **761**, 16-24.

Casali, M., Adamson, A., Alves de Oliveira, C., Almaini, O., Burch, K., Chuter, T. Elliot, J., Forger, M., Foucaud, S., Hambly, N., Hastie, M., Henry, D., Hirst, P., Irwin, M., Ives, D., Lawrence, A., Laidlaw, K., Lee, D., Lewis, J., Lunney, D., McLay, S., Montgomery, D., Pickup, A., Read, M., Rees, N., Robson, I., Sekiguchi, K., Vick, A., Warren, S., and Woodward, B. 2007. The UKRIT wide-field camera. *A&A* **467**, 777-784.

Chambers, J. E. and Wetherill, G. W. 1998. Making the terrestrial planets: N-body integrations of planetary embryos in three dimensions. *Icarus* **136**, 304-327.

Chapman, C. R., Johnson, T. V., and McCord, T. B. 1971. A review of spectrophotometric studies of asteroids. *Proc. Of IAU Colloq.* 12, (Gehrels, T. ed.) Nat. Aero. Space Admin. SP 267.

Chapman, C. R. and Davis, D. R. 1975. Asteroid collisional evolution: Evidence for a much larger early population. *Science*. **190**, 553-556.

Chapman, C.R., Gaffey, M., and McFadden, L., 24-color Asteroid Survey. EAR-A-DBP-3-RDR-24COLOR-V2.1. NASA Planetary Data System, 1993.

Clark, B. E., and Johnson, R. E. 1996. Interplanetary weathering: surface erosion in outer space. *Eos Trans. AGU* **77**, 141-145.

Clark, R. N., Swayze, G. A., Wise, R., Livo, K., Hoefen, T., Kokaly, R., and Sutley, S. 2007. USGS digital spectral library splib06a: U.S. Geological Survey. *Digital Data Series* 231.

Clark, S. P. Jr. 1957. Absorption spectra of some silicates in the visible and near infrared. *Am. Mineral*. **42**, 732-742.

Cloutis, E.A., Gaffey, M.J., Jackowski, T.L. and Reed, K.L. 1986. Calibrations of phase abundance, composition, and particle size distribution of olivine-orthopyroxene mixtures from reflectance spectra. *J. Geophys. Res.* **91**, 11641-11653.

Cloutis E. A. and Gaffey M. J. 1991. Pyroxene spectroscopy revisited: Spectral-compositional correlations and relationships to geothermometry. *J. Geophys. Res*. **91**, 11641-11653.

Cousins A.W.J. 1976. VRI standards in the E regions. *Mem. R. Astr. Soc*. **81**, 25.

DeMeo, F.E., Binzel, R.P., Silvan, S.M., and Bus, S.J. 2009. An extension of the Bus asteroid taxonomy into the near-infrared. *Icarus* **202**, 160-180.

Denevi, B. W., Beck, A. W., Coman, E. I., Thomson, B.J., Ammannito, E., Blewett, D.T., Sunshine, J.M., De Sanctis, M.C., Li J.Y., Marchi, S., Mittlefehldt, D.W., Petro, N.E., Raymond, C.A., and Russell, C.T. 2016. Global variations in regolith properties on asteroid Vesta from Dawn's low-altitude mapping orbit. *Meteorit. & Plan. Sci.* **51**, 2366-2386.

Erasmus, N., Mommert, M., Trilling, D.E., Sickafoose, A.A., van Gend, C., and Hora, J.L. 2017. Characterization of near-earth asteroids using KMTNET-SAAO. *AJ* **154**, 162-171.

Farinella, P., Davis, D. R., Cellino, A., and Zappala, V. 1992. From asteroid clusters to families: A proposal for a new nomenclature. In: Harris, A.W., Lowell, E. (eds.), *Asteroids, Comets, Meteors 1991,* Lunar Planet. Institute, Houston, 165-166.

Fieber-Beyer, S.K., Fieber-Beyer IRTF Mainbelt Asteroid Spectra V3.0. EAR-A-I0046-3-FBIRTFSPEC-V3.0. NASA Planetary Data System, 2015.

Fieber-Beyer, S.K. and Gaffey, M.J. 2019. NEA 162385 (2000 BM19): Near-infrared compositional analysis of a mining target. *Icarus* **328**, 22-31.

Fisher, H. 1941. Farbmessungen an kleinen planeten. *Astron. Nachr* **272**, 127-147.

Fornasier, S., Barucci, M.A., Binzel, R.P., Birlan, M., Fulchignoni, M., Barbieri, C., Bus, S.J., Harris, A.W., Rivkin, A.S., Lazzarin, M., Dotto, E., Michalowski, T., Doressoundiram, A., Bertini, I., and Peixinho, N. 2003. A portrait of 4979 Otawara, target of the Rosetta space mission. *A&A* **398**, 327-333.

Fukugita, M., Ichikawa, T., Gunn, J.E., Doi, M., Shimasaku, K., and Schneider, D.P. 1996. The Sloan Digital Sky Survey photometric system. *AJ* **111**, 1748-1756.

Gaffey, M.J., Burbine, T.H., Piatek, J.L., Reed, K.L., Chaky, D.A., Bell, J.F., and Brown, R.H. 1993. Mineralogical variations within the S-type asteroid class. *Icarus* **106**, 573-602.

Gaffey, M. J. 2010. Space weathering and the interpretation of asteroid reflectance spectra. *Icarus* **209**, 564–574.

Gold T. 1955. The lunar surface. *Mon. Not. R. Astron. Soc*. **115**, 585–604.

Gradie, J., and Tedesco, E. 1982. Compositional structure of the asteroid belt. *Science* **216**, 1405-1407.

Hapke B. 2001. Space weathering from Mercury to the asteroid belt. *J. Geophys. Res*. **106**, 10039–10074.

Hardersen, P.S., Hardersen IRTF Asteroid NIR Reflectance Spectra V1.0. EAR-A-10046-3-HARDERSENSPEC-V1.0. NASA Planetary Data System, 2016.

Hepner, G.F., Logan, T., Ritter, N., and Bryant, N. 1990. Artificial neural network classification using a minimal training set: Comparison to conventional supervised classification. *Photogram. Eng. Remote Sens*. **56**, 469-473.

Hirayama, K. 1918. Group of asteroids probably of common origin. *Proc. Phys. Math. Soc. Japan* **11**, 354-361.

Hiroi, T., Pieters, C.M., and Takeda, H. 1994. Grain size of the surface regolith of asteroid 4 Vesta estimated from its reflectance spectrum in comparison with HED meteorites. *Meteoritics* **29**, 394-396.

Howell, E. S., Merenyi, E., and Lebofsky, L. A. 1994. Classification of asteroid spectra using a neural network. *Journal of Geophysical Research* **99**, 10847-10865.

IAU 2006. Resolution B5: Definition of a planet in the Solar System.
<www.iau.org/static/resolutions/Resolution_GA26-5-6.pdf>. Retrieved March
29, 2019.

IAU 2019. Potentially Hazardous Asteroids (PHAs).
<www.minorplanetcenter.net/iau/Dangerous.html>. Retrieved March 29,
2019.

Johnson, H.L. 1966. Astronomical measurements in the infrared. *ARA&A* **4**, 193-206.

Keller, L.P. and McKay, D.S. 1997. The nature and origin of rims on lunar soil
grains. *Geochimica et Cosmochimica Acta* **61**, 2331-2341.

King T. V. and Ridley W. I. 1987. Relation of the spectroscopic reflectance of olivine
to mineral chemistry and some remote sensing applications. *J. Geophys. Res.*
92, 11457- 11469.

Kitamura, M. 1959. A photoelectric study of colors of asteroids and meteorites. *Pub.
Astron. Soc. Japan*, **11**, 79-89.

Lazzaro, D., Angeli, C.A., Carvano, J.M., Mothe-Diniz, T., Duffard, R., and
Florczak, M.. 2004. S3OS2: The visible spectroscopic survey of 820 asteroids.
Icarus **172**, 179.

Lazzaro, D., Angeli, C.A., Carvano, J.M., Mothe-Diniz, T., Duffard, R., and
Florczak, M., Small Solar System Objects Spectroscopic Survey V1.0. EAR-
A-I0052-8-S3OS2-V1.0. NASA Planetary Data System, 2006.

Lebofsky, L. A., Sykes, M. V., Tedesco, E. F., Veeder, G. J., Matson, D. L., Brown,
R. H., Rudy, R. J. 1986. A refined "standard" thermal model for asteroids
based on observations of 1 Ceres and 2 Pallas. *Icarus* **68**, 239-251.

Lecar, M. and Franklin, F.A. 1973. On the original distribution of the asteroids. I.
Icarus **20**, 422-436.

Merenyi, E., Singer, R.B., and Miller, J.S. 1996. Mapping of spectral variations on
the surface of mars from high spectral resolution telescopic images. *Icarus*
124, 280-295.

Mitchell, T. 1997. *Machine Learning*. New York: McGraw-Hill Inc. 1st ed.

Mizuno, H. 1980. Formation of the giant planets. *Prog. Theor. Phys*. **64**, 544-557.

Mommert, M., Trilling, D.E., Borth, D., Jedicke, R., Butler, N., Reyes-Ruiz, M.,
Pichardo, B., Petersen, E., Axelrod, T., and Moskovitz, N. 2016. First results
from the rapid-response spectrophotometric characterization of near-earth
objects using UKIRT. *AJ* **151**, 98-107.

Moroz, L., Fisenko, A., Semjonova, L., Pieters, C., and Korotaeva, N. 1996. Optical effects of regolith processes on S asteroids as simulated by laser shots on ordinary chondrite and other mafic materials. *Icarus* **122**, 366-382.

Ninomiya, Y. and Sato, I. 1990. Estimation of SiO2 content using thermal infrared reflectance spectra of rocks. *Proceedings IGARSS' 90 Symposium II*, 979.

Noble, S. K., Pieters, C. M., Taylor, L. A., Morris, R.V., Allen, C.C., McKay, D.S., and Keller, L.P. 2001. The optical properties of the finest fraction of lunar soil: Implications for space weathering. *Met. & Planet. Sci.* **36**, 31-42.

Pedregosa, F, Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. 2011. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res*. **12**, 2825-2830.

Pieters, C. M., Taylor, L. A., Noble, S. K., Keller, L.P., Hapke, B., Morris, R.V., Allen, C.C., McKay, D.S., and Wentworth, S. 2000. Space weathering on airless bodies: Resolving a mystery with lunar samples. *Met. & Planet. Sci.* **35**, 1101-1107.

Pieters, C. M. and Noble, S. K. 2016. Space weathering on airless bodies. *JGR-Planets* **121**, 1865-1884.

Pollack, J. B. 1984. Origin and history of the outer planets: Theoretical models and observations L constraints. *ARA&A* **22**, 389-424.

Rayner, J. T., Toomey, D. W., Onaka, P. M., Denault, A.J., Stahlberger, W.E., Vacca, W.D., Cushing, M.C., and Wang, S. 2003. SpeX: A medium-resolution 0.8-5.5 micron spectrograph and imager for the NASA Infrared Telescope Facility. *PASP* **115**, 362-382.

Ramirez, I., Michel, R., Sefako, R., Tucci Maia, M., Schuster, W.J., van Wyk, F., Melendez, J., Casagrande, L., and Castilho, B.V. 2012. The UBV(RI)c colors of the sun. *ApJ* **752**, 5-17.

Reddy, V., Reddy IRTF Near-Earth Asteroid Spectra V1.0. EAR-A-10046-3-REDDYSPEC-V1.0. NASA Planetary Data System, 2010.

Reddy, V. and Sanchez, J.A., Reddy Main Belt Asteroid Spectra V1.0. EAR-A-10046-3-REDDYMBSPEC-V1.0. NASA Planetary Data System, 2016.

Reddy, V. and Sanchez, J.A., Reddy near-Earth and Mars-crossing asteroids V1.0. urn:nasa:pds:ast_spectra_reddy_neos_marscrossers. NASA Planetary Data System, 2017.

Russell C. T., Raymond C. A., Coradini A., McSween H.Y., Zuber M. T., Nathues A., Sanctis M. C. D., Jaumann R., Konopliv A. S., Preusker F., Asmar S. W.,

Park R. S., Gaskell R., Keller H. U., Mottola S., Roatsch T., Scully J. E. C., Smith D. E., Tricarico P., Toplis M. J., Christensen U. R., Feldman W. C., Lawrence D. J., McCoy T. J., Prettyman T. H., Reedy R. C., Sykes M. E., and Titus T.N. 2012. Dawn at Vesta: Testing the protoplanetary paradigm. *Science* **336**, 684-686.

Russell C. T., Raymond C. A., Jaumann R., McSween H. Y., De Sanctis M. C., Nathues A., Prettyman T. H., Ammannito E., Reddy V., Preusker F., O'Brien D. P., Marchi S., Denevi B. W., Buczkowski D. L., Pieters C.M., McCord T. B., Li J.-Y., Mittlefehldt D. W., Combe J.-P., Williams D. A., Hiesinger H., Yingst R. A., Polanskey C. A., and Joy S. P. 2013. Dawn completes its mission at 4 Vesta. *Meteoritics & Planetary Science* **48**, 2076–2089.

Sanchez, J.A., Reddy, V., Kelley, M.S., Cloutis, E.A., Bottke, W.F., Nesvorny, D., Lucas, M.P., Hardersen, P.S., Gaffey, M.J., Abell, P.A. and Le Corre, L. 2014. Olivine-dominated asteroids: Mineralogy and origin. *Icarus* **228**, 288-300.

Sasaki, S., Nakamura, K., Hamabe, Y., Kurahashi, E., and Hirio, T. 2001. Production of iron nanoparticles by laser irradiation in a simulation of lunar-like space weathering. *Nature* **410**, 555-557.

Taylor, L. A., Pieters, C. M., Keller, L. P., Morris, R. V., and McKay, D. S. 2001. Lunar mare soils: Space weathering and the major effects of surface-correlated nanophase Fe. *Journal of Geophysical Research* **106**, 27985-27999.

Tedesco, E. F., Williams, J. G., Matson, D. L., Weeder, G.J., Gradie, J.C., and Lebofsky, L.A. 1989. A three-parameter asteroid taxonomy. *Astron. J.* **97**, 580-606.

Tholen, D.J. 1984. Asteroid taxonomy from cluster analysis of photometry. Ph.D. Thesis, Arizona Univ., Tucson, 1-166.

Usui, F., Hasegawa, S., Ootsubo, T., and Onaka, T. 2019. AKARI/IRC near-infrared asteroid spectroscopic survey: AcuA-spec. *Ast. Soc. Japan* **71**, 1-41.

Xu, S., Binzel, R.P., Burbine, T.H., and Bus, S.J. 1995. Small main-belt asteroid spectroscopic survey: Initial results. *Icarus* **115**, 1-35.

Xu, S., R.P. Binzel, T.H. Burbine, and S.J. Bus, Small Main-Belt Asteroid Survey. EAR-A-M3SPEC-3-RDR-SMASS-V2.1. NASA Planetary Data System, 1996.

Wood, X. H. J. and Kuiper, G. P. 1963. Photometric studies of asteroids. *ApJ* **137**, 1279-1285.

Zellner, B., Tholen, D., and Tedesco, E. 1985. The eight-color asteroid survey: Results for 589 minor planets. *Icarus* **61**, 355-416.