American Mineralogist, Volume 110, pages 685–698, 2025

# Single- and multi-mineral classification using dual-band Raman spectroscopy for planetary surface missions

TIMOTHY K. JOHNSEN[1,2,3,*,†] AND VIRGINIA C. GULICK[1,2,4,‡]

[1]Planetary Systems Branch, NASA Ames Research Center, MS 239-20, Moffett Field, California 94035, U.S.A.
[2]SETI Institute, 339 Bernardo Avenue, Suite 200, Mountain View, California 94043, U.S.A.
[3]Computational Science Research Center, San Diego State University, 5500 Campanile Drive, San Diego, California 92182, U.S.A.
[4]Department of Planetary Sciences, Lunar and Planetary Lab, University of Arizona, 1629 E. University Boulevard, Tucson, Arizona 85721, U.S.A.

## ABSTRACT

Planetary surface missions have greatly benefitted from intelligent systems capable of semi-autonomous navigation and surveying. However, instruments onboard these missions are not similarly equipped with automated science analysis classifiers onboard rovers, which can further improve scientific yield and autonomy. Here, we present both single- and multi-mineral autonomous classifiers integrated using the results from a co-registered dual-band Raman spectrometer. This instrument consecutively irradiates the same spot size on the same sample using two excitation lasers of different wavelengths (532 and 785 nm). We identify the presence of mineral groups: pyroxene, olivine, potassium feldspar, quartz, mica, gypsum, and plagioclase, in 191 rocks. These minerals are among the major rock-forming mineral groups, so their presence or absence within a sample is key for understanding rock composition and the environment in which it formed. We present machine learning methods used to train classifiers and leverage the multiple modalities of the dual-band Raman spectrometer. When testing on a novel sample set for single-mineral classification, we show accuracy scores up to 100% (varying by mineral), with a total classification rate (all minerals) of 91%. When testing on a novel set of samples for multi-mineral classification, we show accuracy scores up to 96%, with a total classification rate of 73%. We end with several hypothesis tests demonstrating that dual-band Raman spectroscopy is more robust and improves the scientific yield for mineral classification over single-band spectroscopy, especially when combined with our multimodal neural network.

**Keywords:** Raman spectroscopy, mineralogy, autonomous classifier, multimodal machine learning, dual-band Raman spectroscopy, sensor fusion

## INTRODUCTION

Future lunar and martian rovers and lunar-suited astronauts aim to navigate terrain using multimodal sensors and Intelligent Data Understanding (IDU) software. As rovers are equipped with more robust instruments and greater navigational autonomy (Francis et al. 2017; Verma et al. 2023), there is a subsequent need to develop systems capable of in situ data analysis for onboard decision making. Raman spectrometers, like Mars 2020s Scanning Habitable Environments with Raman and Luminescence for Organics and Chemicals (SHERLOC) (Beegle et al. 2015), SuperCam (Maurice et al. 2021; Wiens et al. 2021), and ExoMars' Raman Laser Spectrometer (RLS) (Rull et al. 2017) have been deployed on Mars rovers to study mineral compositions and detect organics. However, none of these instruments possess onboard capabilities to autonomously identify mineral, organic, or rock compositions. Instead, all data must be returned to Earth and await confirmation for identified minerals and commands on what to do next. This creates a significant bottleneck in scientific decision making, the command cycle, and subsequent traverse planning. To further improve data collection, a growing trend is to use information from multiple spectral techniques (for example, Raman, NIR, Mid-IR, XRD, and LIBS) and imaging sources to analyze rock and sediment samples on planetary surfaces. Our research aims to demonstrate that machine learning methods that combine multi-laser excitation Raman spectra to classify minerals in natural rock and sediment samples represent a step forward in automated spectral classification for single- and multi-mineral samples. Furthermore, we show end-to-end methods for model selection and error mitigation.

Applications of Raman spectroscopy to mineral identification have paved the way for autonomous mineral classification of both single- and multi-mineral mixtures using Raman spectra and machine learning. However, there are three knowledge gaps that we will address: (1) the use of single-band Raman spectrometers vs. dual-band Raman spectrometers; (2) the use of either synthetic mineral mixtures to train models or a small variety of samples to test solutions to the multi-mineral mixture problem; and (3) the use of more effective machine learning methods with which to leverage dual-band Raman spectra. We also compare key findings of previous studies to motivate our classifier development with the current results.

We present a machine learning pipeline that inputs Raman spectra from natural samples and outputs the likelihoods of

* Corresponding author E-mail: tim.k.johnsen@gmail.com, Orcid https://orcid.org/0000-0002-0698-3875
† Open access: Article available to all readers online.
‡ Orcid https://orcid.org/0000-0002-8181-3112

0003-004X/25/0005–685$05.00/DOI: https://doi.org/10.2138/am-2023-9072

various minerals present. We demonstrate that using a pipeline of methods combining our dual-band Raman spectrometer, data augmentation, averaging, normalization, early stopping (Prechelt 1998), artificial neural networks (ANN), and multimodal deep learning (Ngiam et al. 2011)—in totality what we call a multimodal neural network (MNN)—is superior to other machine learning approaches using either linear regression or a multi-layer perceptron and those which utilize only a single-band Raman spectrometer. In the next section, we provide the background pertinent to understanding the history, key findings, parameters, and knowledge gaps we address with our study. In the "Methods" section, we discuss methods used in our pipeline from model input to output, which includes spectra preprocessing, classical machine learning algorithms, and our multimodal neural network. We evaluate our models in the "Results" section, demonstrating the advantages and robustness of our MNN approach, resulting in a mean pure-mineral classification rate of 91% and multi-mineral classification rate of 73%—each accuracy varies by mineral. A series of hypothesis testing that quantifies the statistical advantage of our MNN approach over classical approaches is included.

## Background

### Previous studies

Raman spectroscopy is a rapid, non-invasive spectral approach to analyze minerals and potential biosignatures in natural rock and sediment geological samples, providing a unique fingerprint of the material (Lewis and Edwards 2001). Autonomous characterization of materials with Raman spectroscopy was first used to discriminate various narcotics using a single-band 785 nm Raman spectrometer, principal component analysis (PCA), and machine learning (Ryder 2002; Howley et al. 2005). Similar techniques have since been used to identify and grade prostatic adenocarcinoma cell lines (Crow et al. 2005) and characterize bacterial species (Xie et al. 2005). Methods used for autonomous single-mineral classification, considering samples as homogenous mixtures, with Raman spectroscopy and machine learning have also been studied.

Freeman et al. (2008) presented a first derivative least-square technique to characterize several feldspar end-members with 94% accuracy, using a single-band 532 nm Raman and the wavenumber range 150 to 1800 $cm^{-1}$, noting that the range 2500 to 4000 $cm^{-1}$ can be important for characterizing hydrated minerals. Wang et al. (2006) concluded that most diagnostic mineral bands are between 400 and 1300 $cm^{-1}$ when using a 532 nm excitation laser to analyze anhydrous and Mg-sulfates. Ishikawa and Gulick (2013) used an 852 nm laser excitation Raman spectrometer of 130 spectra from their mineral library. Spectra from samples were obtained without ambient light and the need for preparation, such as grinding or pulverizing. Therefore, this approach enables the analysis of samples in situ without moving or altering them. They then tested the robustness of their classifier by using mineral spectra from the RRUFF library (Downs 2006), which were measured using different Raman spectrometers with varying single-band excitation lasers from 514 to 852 nm. Ishikawa and Gulick's automated mineral classifier was able to discriminate the six minerals tested: mica, olivine, plagioclase, potassium feldspar, pyroxene, and quartz, with up to 100% accuracy (varying by mineral) and 83% total accuracy (among all minerals). They found an Artificial Neural Network (ANN) to be more robust than a decision tree and that including the background fluorescence patterns improves classification accuracy. Before this study, background fluorescence was generally viewed as background noise and typically removed, while only diagnostic peaks were used for identification.

To address the multi-mineral classification of heterogenous mixtures, Lopez-Reyes et al. (2014) used multivariate analysis models based on PCA, partial least-squares, and ANNs to show a relationship between ANN output and mineral abundancies of selected Ca-, Fe-, Na-, and Mg-sulfates in binary mixtures. They used a laboratory version of the RLS aboard the 2018 ExoMars mission (Rull et al. 2017), which uses a 532 nm excitation laser. Their results showed that an ANN can distinguish 17 sulfides with 100% accuracy. Any of those sulfates could be detected from a simulated linear combination of their spectra with 100% accuracy if at least 10% of that spectrum was present in the linear combination. Cochrane and Blacksberg (2015) presented a machine learning approach to find an optimal linear combination of pure mineral spectra to recreate a mixture of minerals with an F1-score of 82% on a synthetic data set created from linear combinations of pure mineral spectra. Berlanga et al. (2022) used a single-band 532 nm Raman spectrometer and a Convolutional Neural Network (CNN) with 60 million trainable parameters to classify if the minerals quartz, albite, microcline, biotite, and hornblende are present on both a slab of granite and gabbro with an F1-score of 99%. They found that these minerals could be identified with an F1-score of 80–81% when half of the surface was covered in dust.

In this study, we improve autonomous multi-mineral classification with Raman spectroscopy and machine learning by: (1) leveraging a dual-band co-registered Raman spectrometer, where previous approaches have only used single-band Raman spectrometers; (2) using our library of dual-band Raman spectra measured from 191 different rocks as a test set, instead of approaches which used either linear combinations of pure mineral spectra to simulate mixtures, several spectra of the same samples, or a relatively small sample size; and (3) presenting a robust ablation study, which is a method of experimentation that incrementally removes/replaces individual parts of the classifier pipeline to evaluate their contribution. The results of the ablation study warrant the benefits of more complex, modern, and novel neural network methods while maintaining a low computation and memory resource overhead used during inference. Such methods can be adopted by remote robotics for improved space exploration and scientific yield, which is evident by a series of hypotheses testing that compared our contributions to less sophisticated, more antiquated approaches.

### Dual-band Raman spectroscopy

Here, we introduce the co-registered, dual-band Raman spectrometer we used, as developed by Spectra Solutions, Inc. (SSI). The term "dual-band" refers to using two lasers at different excitation wavelengths. The term "co-registered" refers to consecutively irradiating the same spot on a sample. Thus, the SSI instrument we used consecutively irradiates the same spot on the same sample with two different excitation wavelengths

(532 and 785 nm). The main benefits of such an instrument are: (1) wavenumber bands of low signal-to-noise ratios in one laser excitation can be supplemented by high signal-to-noise ratios in another; (2) there is an intrinsic difference in background fluorescence between the two excitation lasers that can be leveraged by a classifier; and (3) each laser can explore a different wavenumber space and resolution. The first excitation laser has a wavelength of 532 nm and a wavenumber range of 130–4000 $cm^{-1}$ at a resolution of 2.5 $cm^{-1}$. The second excitation laser has a wavelength of 785 nm and a wavenumber range of 190–2900 $cm^{-1}$ at a resolution of 0.5 $cm^{-1}$. Both excitation lasers utilize a spot size of ~50 μm.

### Neural networks

For deployment, we developed source code in raw C++, with no external dependencies; we demonstrated its functionality in a previous paper (Johnsen et al. 2020). Python was used in this study for easy visualization, access to libraries, portability, and model development. We used the Python library Keras (Chollet 2015) with TensorFlow (Abadi et al. 2015) for some neural network implementations.

Each of our models is a form of an Artificial Neural Network (ANN), where we input a spectrum that gets fed through the ANN to learn latent features, which are used to output likelihoods of each mineral being present in the spectrum and a relative measurement of confidence. We evaluate two classical models for mineral classification: a Logistic Regression (LR), such as that used by Cochrane and Blacksberg (2015), and a Multi-Layer Perceptron (MLP), such as that used by Ishikawa and Gulick (2013). LR takes a linear combination of the input spectrum and passes it through the Sigmoid function, which is bounded between 0 and 1. Two key benefits of LR are its "explainability" due to the simple weights learned during training and the fact that few model parameters needed to be executed during run time. The disadvantage of LR is that it may not be complex enough to learn the nonlinear relationship between the spectrum and present minerals. An MLP can have a higher complexity to facilitate learning more complex relationships but comes at the cost of losing explainability and increasing computing and memory overhead. We compare an LR approach to an MLP to see if the added overhead and loss in explainability are worth the potential increase in accuracy. Note that the output of an ANN is an arbitrary value between 0 and 1 and should not be interpreted as a statistical probability or percentage of the mineral being present in the sample. Instead, it is more of a confidence score, where a value closer to 1 means there is a higher confidence that a mineral is present than a value closer to 0.

We then compare a third approach using an evolved form of Multimodal Deep Learning (MDL) (Ngiam et al. 2011). Ngiam introduced MDL with an example that combined both audio and video modalities for speech recognition. Restricted Boltzmann machines (RBM) (Salakhutdinov 2007) refer to an architecture that trains neural networks without needing to know ground truth labels for input data (unsupervised learning). RBMs are used to learn single modality features independently and are then used as feature extractors to input into a downstream neural network that captures cross-modality features used in classification—using the known ground truth labels (supervised learning). We update MDL with more state-of-the-art processes and specific methods for our application, creating what we call a multimodal neural network (MNN), as detailed in the "Methods: Multimodal neural network" section.

## Methods

### Data collection

There are three data sets of samples for our studied Raman spectra: pure mineral samples measured using our dual-band co-registered SSI Raman spectrometer, multi-mineral rock samples measured using the same SSI spectrometer, and pure mineral samples whose spectra were collected from several labs across the world and aggregated on the publicly available online database RRUFF (Ambruster and Danisi 2015). Table 1 lists the number of spectra and samples from each data set, along with identified minerals in the samples of which the spectra were measured. Figure 1 shows all our pure mineral spectra plotted together and obtained by both excitation lasers with the SSI dual-band, co-registered Raman spectrometer. Online Materials[1] Appendix A contains Tables A1, A2, and A3, which list a mixture of the following information for each sample in our data sets: their ground truth mineral compositions and uncertainties resulting from different geologists, the locations where the sample was collected, the person who collected the sample, and the minerals identified by the neural network along with the corresponding confidence score output from the neural network.

Many of our mineral samples came from rock kits from DJ Minerals (see Online Materials[1] Appendix A for more details). These samples were from Madagascar, Canada (Ontario), the U.S.A. (Montana, Wyoming, Arizona, South Dakota, Fremont County, Colorado, and Eddy County), and Mexico. For multi-mineral (rock) samples, we collected rocks from the U.S.A. (Arizona, California, Utah, Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, South Dakota, Washington, and Wyoming) and Chile. We used 191 of our rock samples in the algorithm development effort, mostly from rock types that we would expect to see on the surface of Mars or the Moon. We broke off one side of each sample with a rock hammer to reveal a fresh surface and to simulate the collection of field samples on a planetary or lunar surface. We measured several Raman spectra on the fresh side of samples containing either a single mineral or multiple minerals. We obtained 966 spectra in total since there were several samples with ground truth positives for more than one mineral. We acquired several spectra from distinct points along transects on the surface of each mineral or rock sample to ensure a representative sample of the mineralogy. We obtained spectra manually under controlled conditions, where each sample was shielded from ambient light by placing it in a darkened enclosure while acquiring spectral pairs. During the acquisition process, we initially put the probe directly above the sample at a distance of approximately 1 mm from the surface and slowly raised the stage until the signal strength was at its strongest (~1–2 cm), roughly coinciding with the focal length of the laser beam.

We supplemented our SSI data set with the public database RRUFF (Ambruster and Danisi 2015). The RRUFF data set is a collection of spectra from multiple labs using different Raman spectrometers with either 532, 780, or 785 nm wavelengths. We used only the spectra in the RRUFF set labeled as homogenous (i.e., pure mineral). Recall that an SSI spectral pair is measured from the same spot on a sample, but the spectra from the RRUFF set are unlikely to be measured from the same spot. We make a necessary assumption that two spectra measured from different wavelengths of the same sample in the RRUFF data set are approximately equivalent to an SSI spectral pair. This assumption is critical when combining and comparing the two data sets. We also assume that spectra obtained with the 780 and 785 nm excitation wavelengths in the RRUFF set are approximately equivalent. We note that the majority of the spectra from the RRUFF database have confirmed mineral compositions from X-ray diffraction and chemical analysis. These assumptions are not ideal (Dyar et al. 2016); however, not only do our results show

**Table 1.** Raman spectral library

| | SSI multi-mineral | | SSI pure-mineral | | RRUFF pure-mineral | |
|---|---|---|---|---|---|---|
| Mineral | Samples | Spectra | Samples | Spectra | Samples | Spectra |
| Gypsum | 4 | 70 | 1 | 3 | 2 | 2 |
| Mica | 138 | 684 | 3 | 11 | 10 | 10 |
| Olivine | 5 | 45 | 2 | 5 | 17 | 17 |
| Plagioclase | 185 | 927 | 5 | 47 | 16 | 16 |
| K-feldspar | 153 | 767 | 3 | 14 | 10 | 10 |
| Pyroxene | 72 | 270 | 4 | 33 | 20 | 20 |
| Quartz | 184 | 918 | 4 | 18 | 10 | 10 |

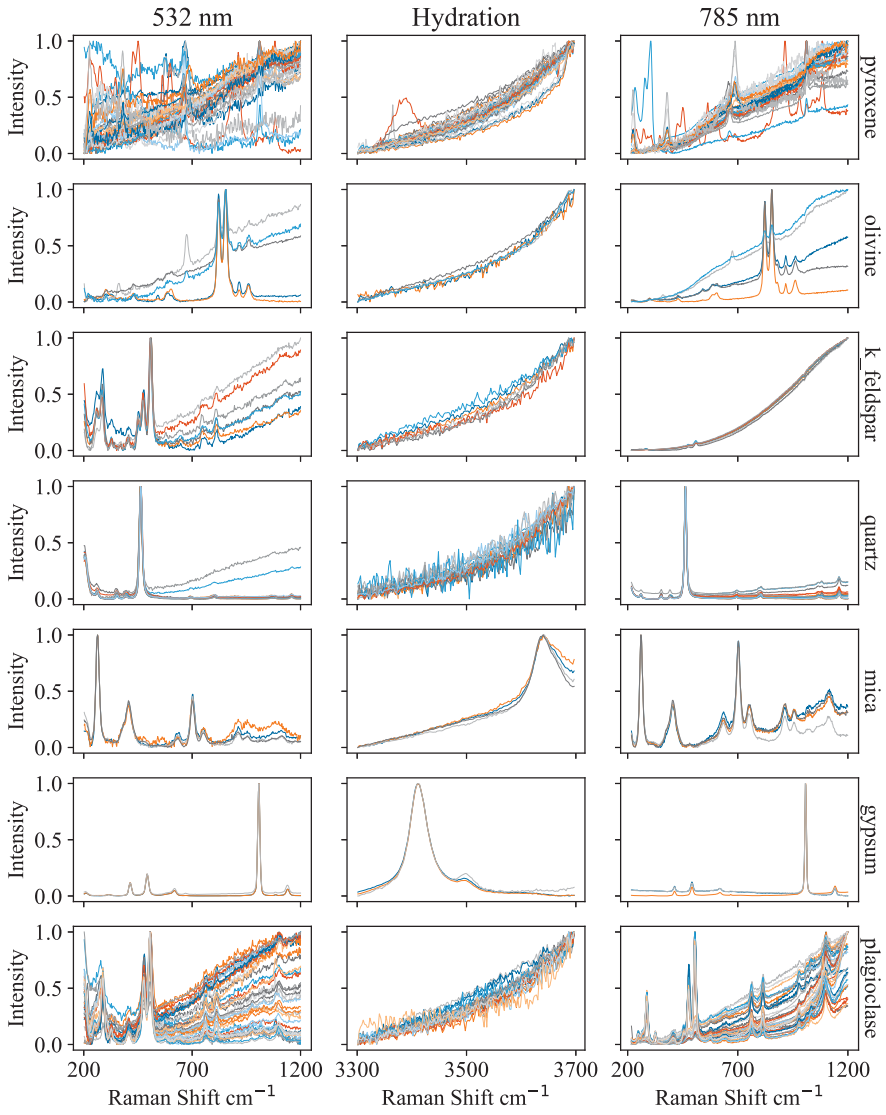


**Figure 1.** Spectra measured using our SSI dual-band Raman spectrometer and pure-mineral samples. The hydration range is also from the 532 nm excitation laser.

that the error resulting from these assumptions is acceptable, but our dual-band data set is the first of its kind, so no other such data sets exist yet in literature. Thus, the advantages of having an arbitrary data set to evaluate models outweigh the disadvantages of these assumptions.

We trained the neural networks on spectra collected from our in-house pure minerals only, in which their mineral compositions were confirmed by comparing their labels provided by the vendor to the spectral bands we measured with our Raman spectrometer and mineral bands from the literature. After training, we evaluated the neural networks on two data sets: (1) the RRUFF spectra and (2) the spectra collected from our in-house multi-mineral rock samples. Our rock samples had their mineral compositions determined by two to three analysts trained in petrology and mineralogy, as corroborated through a combination of techniques: visual inspection, hand sample analysis, thin section analysis, labels provided by the mineral suppliers, the geological context of the areas in which the samples were collected, and comparing Raman spectral bands of our samples to those both in the RRUFF data set and in the literature. Since the ground truth compositions of the rocks were independently derived from the results of three hand sample analysts, the overlapping range of percentages may be less than 100%, as shown in Online Materials[1] Table A3. Note that the classifier only looks for the presence or absence of a mineral, regardless of the percentage of that mineral present in the rock. Online Materials[1] Table A3 also includes which minerals were detected from the final classifier, along with the confidence score output from the ANN.

## Data processing

To reduce the spectral ranges to those containing diagnostic spectral peaks of each mineral being investigated, we first truncated wavenumber ranges commonly within minerals that have low signal-to-noise ratios. We determined that the range 200 to 1200 cm$^{-1}$ is sufficient for minerals in our study, based on both previous literature results (Wang et al. 1995, 2006; Freeman et al. 2008) and personal observations. For example, Wang et al. (1995) reported that the principal Raman peaks are for olivine at ~956 and ~826 cm$^{-1}$; for plagioclase at 503 and 483 cm$^{-1}$; for quartz at 465, 207, and 128 cm$^{-1}$; for potassium feldspar at 513, 475, and 454 cm$^{-1}$; orthopyroxene at 1006, 678, and 660 cm$^{-1}$; and for clinopyroxene at 1005 and 665 cm$^{-1}$. We also include the wavenumber range 3300 to 3700 cm$^{-1}$ because this range shows hydration bands crucial to distinguishing some minerals [like gypsum 3500 cm$^{-1}$ (Liu et al. 2009)] and organics. This yields two main modalities: A, the 532 nm excitation laser using the two wavenumber ranges 200–1200 cm$^{-1}$ ($A_1$) and 3300–3700 cm$^{-1}$ ($A_2$), and B, the 785 nm excitation laser using the wavenumber range 200–1200 cm$^{-1}$.

Spectra are collected into matrix $\mathbf{X}$, where each row is a spectrum, and each column is the intensity corresponding to a wavenumber. We normalized the intensity values ($\mathbf{X}$) of each spectrum by their respective minimum *min* ($\mathbf{X}_{j,*}$) and maximum intensity *max* ($\mathbf{X}_{i,*}$), using Equation 1.

$$\mathbf{X}_{i,*} = \left[\mathbf{X}_{i,*} - min(\mathbf{X}_{i,*})\right] / \left[max(\mathbf{X}_{i,*}) - min(\mathbf{X}_{i,*})\right] \quad (1)$$

Normalization is a critical step in preprocessing because the scale of intensity values between different spectra can vary widely with integration time, orientation, expected composition, working distance, environment, and spectrometer. Dyar et al. (2016) showed that normalization may not fully account for such phenomena; however, we find it empirically adequate for use within our machine learning models, especially after wavenumber truncation (Fig. 1). Furthermore, we empirically find that by introducing Gaussian noise into the training process, as shown by Bengio et al. (2014), we can capture expected errors and inherent noise in data and thus do not need to employ more evasive processing techniques. There is a tradeoff when using smoothing and baseline removal techniques. The main benefit is more prominent diagnostic peaks and less background fluorescence, whereas the disadvantages are the inherent destruction of original information and the creation of artifacts. Both Ishikawa and Gulick (2013) and this study noted improvement in classification accuracy using normalized but otherwise raw spectra, as opposed to processed spectra. The complex relationship between machine learning classification and various processing techniques requires more study.

The two processing techniques that use wavenumber truncation and intensity normalization are used with all our models because the models require a certain level of dimensionality reduction and scale adjustment. However, two processing techniques are not required, so we turn them on and off to compare performance between the machine learning models: averaging multiple spectra from the same sample and PCA.

For multi-mineral rock samples, we compare taking the average of all spectra for an entire sample. Theoretically, this may mitigate variance and dependence on the location of samples used to collect a spectrum. However, this action may also mask diagnostic peaks needed to classify certain biosignatures and minerals. We also compare using PCA to compact the data further and highlight inherent variance. Note that improved accuracy can be obtained if the training and validation/test sets are combined into the data used to calculate the basis for PCA. However, because the testing set is the actual data collected from a novel environment, we only calculated the basis for PCA using the training set. During an actual surface mission, calibration may include field samples in the data used to calculate the basis for PCA; however, the models will then need to be retrained. After calculating the basis for PCA, we use only the principal components that contribute at least 1% to the total variance. This reduces the dimensions from $10^3$ to $10^1$, where the exact dimensions vary depending on the training set and modality being used. After PCA, the columns of intensity values ($\mathbf{X}$) are standardized to zero mean and unit variance using Equation 2.

$$\mathbf{X}_{*_j} = \left[\mathbf{X}_{*_j} - mean(\mathbf{X}_{*_j})\right] / \left[standard\_deviation(\mathbf{X}_{*_j})\right] \quad (2)$$

## Data augmentation

We also compare a method for improving training data called data augmentation. We use data augmentation to: (1) simulate mineral mixtures and (2) balance the unequal distribution of spectra used in each mineral class, as seen in Table 1. This process generates additional training data where there is insufficient spectral data (Cochrane and Blacksberg 2015). Spectral data are augmented by taking random linear combinations of 1–4 pure mineral spectra from our SSI data set. Note that there is a nonlinear relationship between mineral mixing and Raman spectroscopy; however, this exact relationship is unknown. Thus, we add Gaussian noise to adjust for this unknown error. This process of augmentation with linear combinations and Gaussian noise is only used during model training, not during inference or in our test sets, and is a stopgap until either: (1) the relationship of mixing minerals, in the context of Raman spectroscopy, is better understood; and/or (2) a larger database of Raman spectra measured from mineral mixtures with known abundancies is available. Figure 2 shows an example of an augmented spectrum compared to a natural spectrum with a similar composition. The standard deviation used to randomize Gaussian noise was empirically selected to be 10% of the spectrum's total mean intensity. Augmentation is carried out during each iteration of model training to create additional unique spectra to balance an equal number of natural spectra drawn from each mineral group. We emphasize that no augmented spectra are ever used in the test set, and all results are reported as tested against spectra from natural samples. We later show that such data augmentations enhance the classification accuracies of our models by allowing them to infer multi-mineral classifications even when provided with only pure mineral spectra.

## Multimodal neural network

A unique challenge is how to best represent the multiple modalities of the dual-band Raman spectrometer, such that they map to the presence of minerals in mixtures. Figure 3 illustrates the three machine learning approaches that we compare. Modality A is spectra obtained from the 532 nm excitation laser, and modality B is from the 785 nm excitation laser. Modality $A_1$ is the first part of modality A (wavenumbers 200–1200 $cm^{-1}$) that corresponds to a spectral region that is rich in mineral information, and $A_2$ is the second part that corresponds to the hydration range (wavenumbers 3300–3700 $cm^{-1}$). One approach would be to concatenate the spectra from modalities A and B and proceed with training the given classifier as if the dual-band Raman spectral data are only from one modality. This is how our vanilla MLP and LR models input data (MLP and LR models have been used in previous literature, as listed in the "Background: Previous studies" section, on single-band spectra). However, training a neural network on spectral data that is concatenated in this way results in hidden layer nodes with higher magnitude weights assigned to individual modality features as opposed to interconnecting weights between different modalities (Ngiam et al. 2011). This is why Ngiam proposed Multimodal Deep Learning (MDL) as a two-step training routine that trains the first part of the neural network on each modality independently and then the second part of the neural network on both modalities. In Online Materials[1] Appendix B, we present our approach to MDL, which consists of the architecture of the neural network followed by our training procedure for such a network. We refer to this neural network as a multimodal neural network (MNN), which is a core contribution of this paper. The main novelties of the MNN over classical MDL approaches are: (1) applying to dual-band spectra as opposed to audio and video modalities, and (2) the training procedure of the neural network, which is more detailed in Online Materials[1] Appendix B.

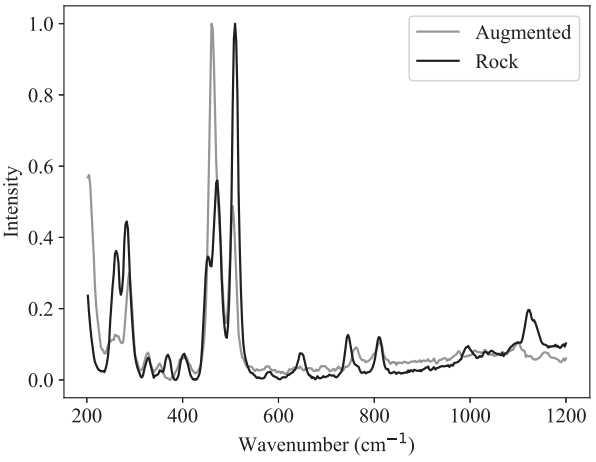


**Figure 2.** A natural rock spectrum compared to an augmented one with similar composition. Augmented spectra are used to supplement the training set and are never used in the testing set.

## Cutoff values and AUC

During the classification process, an input spectrum will receive a given likelihood from the ANN (between 0 and 1) for the presence of each mineral. A threshold, otherwise called the "cutoff" value, can be set to weed out bad results. Theoretically, samples with a lower percent composition should have lower likelihoods. As the cutoff value increases, the number of false positives decreases but the number of true positives also decreases. A known cost function must be used to optimize the cutoff value, assigning a cost to false positives and true positives.

In our classification scheme, ground truth positive samples should have higher likelihoods than ground truth negative samples, and ideally, there should be a definitive line separating the two classes. This "separability" is measured using the Area Under the receiving operating characteristic Curve (AUC). AUC measures the separation by iteratively sliding the cutoff values between 0 and 1 and measuring the true positive rate against the false positive rate. An AUC score of 1 reflects a classifier with perfect separation between the classes, a value of 0.5 reflects a purely random binary classifier, and a value below 0.5 reflects separation worse than a random binary classifier. The goal is to have some intermediate value between 0.5 and 1, where the closer to 1, the better the separation, resulting in a more accurate classifier, regardless of the cutoff value. During an actual surface mission or given a specific task, the user will determine cutoff values depending on a well-defined cost function, in which quantities such as accuracy will have more significance. Without a defined cost function, our classifiers are designed to maximize AUC scores since they are highly correlated to other error metrics without defining a specific cutoff value. For demonstration purposes, we selected cutoff values by optimizing a metric called "balanced

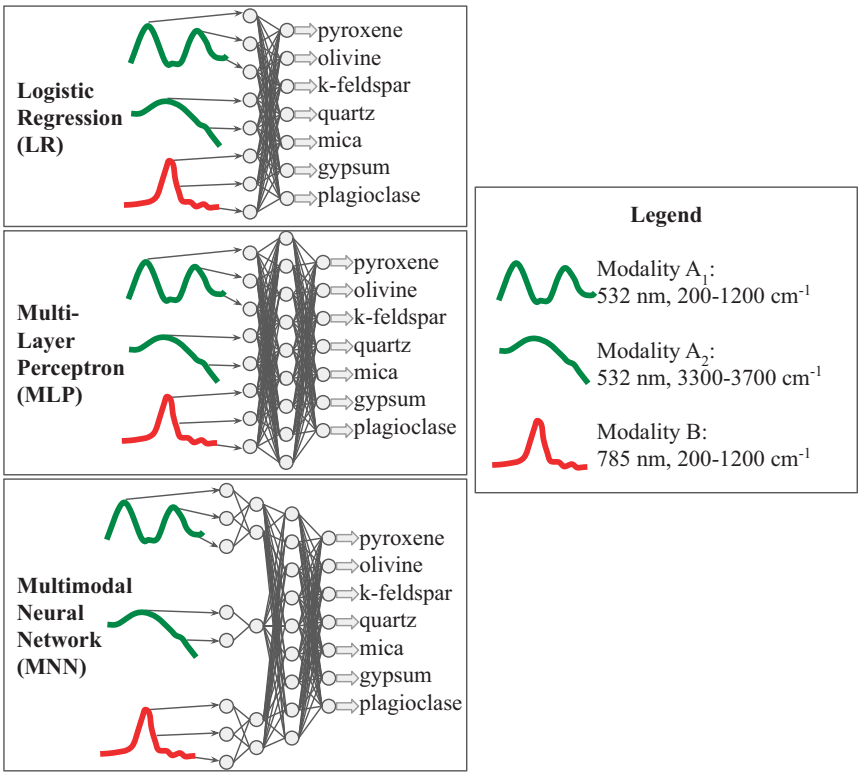


**Figure 3.** Illustration of the neural architectures for the three machine learning approaches. Each approach increases in complexity, going from LR to MLP to MNN. The first input layer directly inputs intensity values at each wavenumber into each node (circles). The lines in between circles show which nodes are connected. Note that the MNN has "stems," so some nodes are not connected to all other nodes in the previous layer (such as in an MLP).

accuracy," which is the average recall of each class, to provide an example of how to make classifications in the field.

## Experimental setup and model selection

We designed a series of experiments to compare several models. We varied: (1) the preprocessing techniques comparing raw spectra vs. averaging and/or PCA; (2) the input modalities comparing Raman excitation wavelengths, 532 or 785 nm, as independent single-band modalities vs. combined dual-band; (3) the complexity of the neural network algorithm comparing LR vs. MLP vs. MNN; and (4) classifying the sample as either a single- or multi-mineral sample considering it as having either a homogeneous or heterogeneous mineral composition, respectively. All model configurations are listed in Table 2. Using different random seeds, we independently trained and evaluated 100 neural networks corresponding to each model. This was done to capture inherent variance in the algorithms, where the number 100 was determined empirically during preliminary testing to capture variance properly.

For each model, we used our SSI Raman spectra measured from 22 pure mineral samples as the training set. The training set contains data used to optimize neural network parameters using a local search algorithm, as detailed in Online Materials[1] Appendix B. Note that we also toggled using data augmentation on these 22 samples during the training process to measure its effect on AUC (separability). After training, we evaluated each model on a holdout testing set, which was not shown to the model during training. We used the RRUFF Raman spectra measured from 85 pure mineral samples as the testing set for single-mineral classification. For multi-mineral classification, we used our SSI Raman spectra measured from 191 multi-mineral rocks as the testing set.

We conducted a series of one-sided hypothesis tests to compare each model. We compared two distribution statistics: the mean AUC score and the standard deviation, σ, in AUC scores across all 100 random runs for each model. A p-value reflects the statistical relevance between the difference of two means, where a low p-value shows a low likelihood that one mean just happens to be higher than the other. A p-value near 0 warrants that a hypothesis is true, and a p-value near 1 does not warrant that a hypothesis is true. We used p-values to compare the AUC scores between two models, to test the hypothesis that one model is more robust than the other.

After comparing models, we selected a final model with a well-shaped AUC distribution across all 100 random runs that had a high mean and low variance and significant p-values warranting statistical relevance that the model was more robust than the others. We analyzed learning curves to select a final trained neural network from the 100 random runs from the selected model. A learning curve visualizes the error vs. iteration of the local search algorithm that is used to optimize the neural network parameters. An ideal learning curve shows smooth convergence to an optimum in the loss function. Note that each neural network converges to a local optimum, which is likely not the global optimum. This phenomenon is the main source of variance between multiple neural networks trained with different random seeds. After selecting a final model and trained neural network, we set the cutoff values used to make final classifications.

The single-mineral classifier can be used to identify a pure mineral sample. Thus, we only needed to make one classification, and we classified the sample as the mineral group with the highest output value from the neural network. The multi-mineral classifier assumes the sample may have a mixed composition. Thus, we made several classifications based on which ones received a neural network output higher than the respective cutoff value for that mineral, resulting in zero to many positive classifications. Since we do not have a defined cost function that is mission specific, see the "Methods: Cutoff values and AUC" section, we selected cutoff values that maximized the balanced accuracy.

Furthermore, we used leave-one-out cross-validation to measure the robustness of the 22 pure mineral samples in our training data set. This process consists of iteratively removing one spectrum from the training set, training a model on the other spectra, testing the one spectrum removed, then placing that spectrum back into the training set and repeating the process until each spectrum is removed exactly once.

## Results

Figures 4–6. show distributions of AUC scores for each trained model across all 100 random runs. A wider distribution corresponds to a larger bin count, showing higher clustering. In contrast, a narrower distribution shows that the AUC scores were more dispersed. A taller distribution shows a higher range of AUC scores. The black horizontal line in each distribution indicates the average AUC score over all 100 runs.

Table 3 shows the statistics and resulting p-values from the one-sided hypothesis tests. The first column in Table 3 shows

**Table 2.** Experimental configurations

| Configuration | Modalities | Processing | Test | Configuration | Modalities | Processing | Test |
|---|---|---|---|---|---|---|---|
| **I** | A | raw | RRUFF | **VII** | A | raw | SSI |
| **II** | B | raw | RRUFF | **VIII** | B | raw | SSI |
| **III** | A, B | raw | RRUFF | **IX** | A, B | raw | SSI |
| **IV** | A | PCA | RRUFF | **X** | A | PCA | SSI |
| **V** | B | PCA | RRUFF | **XI** | B | PCA | SSI |
| **VI** | A, B | PCA | RRUFF | **XII** | A, B | PCA | SSI |
| | | | | **XIII** | A | mean | SSI |
| | | | | **XIV** | B | mean | SSI |
| | | | | **XV** | A, B | mean | SSI |
| | | | | **XVI** | A | mean, PCA | SSI |
| | | | | **XVII** | B | mean, PCA | SSI |
| | | | | **XVIII** | A, B | mean, PCA | SSI |

the methods being compared. Mode A refers to the 532 nm Raman excitation laser and mode B to the 785 nm; augmented refers to adding augmented spectra to the training set (not the testing set); logistic regression (LR) is the lowest complexity machine learning algorithm (done before in literature), multi-layer perceptron (MLP) is the next highest complexity (done before in literature). Multimodal neural networks (MNNs) have the highest complexity (introduced in this paper for mineral classification). The second column indicates either the single- or multi-mineral classifier. The third column shows the p-value from a T-test given the alternative hypothesis, $H_a$, where the null hypothesis is the inverse of $H_a$. A p-value near 0 indicates strong statistical relevance that method 1 is better than method 2, whereas a p-value near 1 indicates a strong statistical relevance that method 1 is not better than method 2. A p-value is reported for both single- and multi-mineral classification. The last four columns show the means and standard deviations of the AUC distributions. Note that when comparing MNN to either MLP or LR, only the set of results that used both modalities A and B were considered——since MNN operates under the assumption that the model inputs multiple modalities. Also, we only compare averaging spectra for multi-mineral classification because the rocks have a much higher variance than pure mineral samples, and thus, averaging is not needed for pure mineral samples.

To select the most robust model and trained neural network, we compare the p-values in Table 3, the AUC distributions shown in Figures 4–6, and the learning curves for each configuration and set of runs. Figure 7 shows the learning curves for all 100 runs from the two selected configurations. Not all the learning curves are shown for each model, for brevity, since there are 8400 of them in total. The final selected model, which had the most robust AUC scores for the single-mineral classifier, uses an MNN, dual-band excitation lasers, and augmented spectra (model configuration III). The final selected model, which had the most robust AUC scores for the multi-mineral classifier,

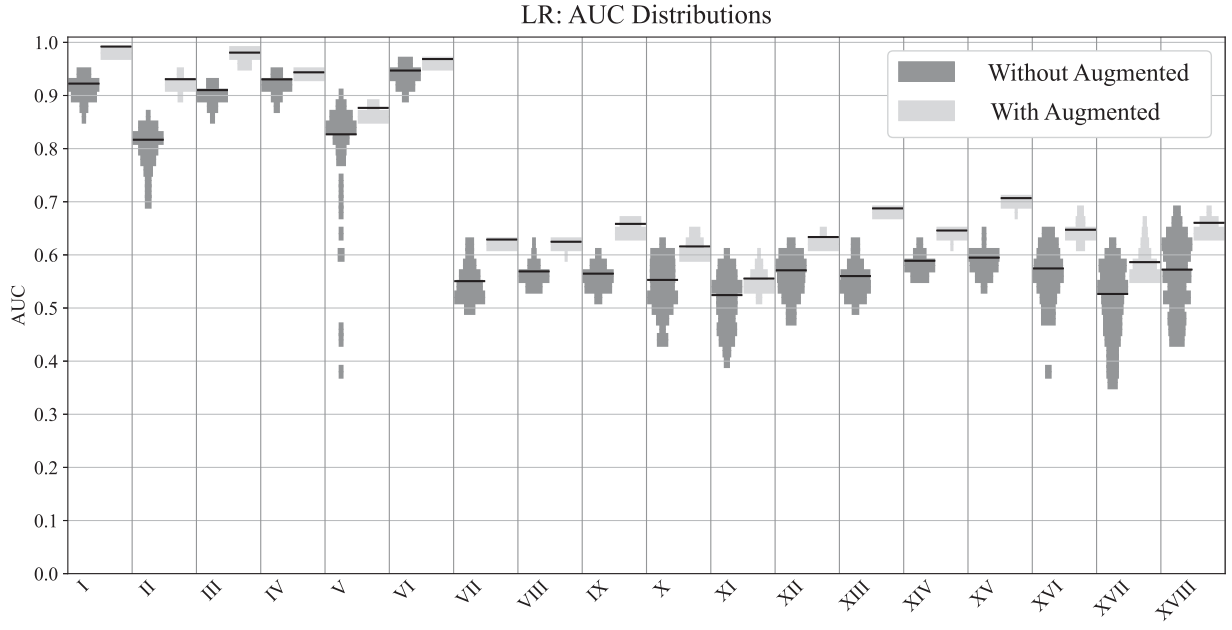


**Figure 4.** Results of each configuration using a Logistic Regression (LR) model. Not using augmented spectra in the training process corresponds to the left-sided distribution in each column and using augmented to the right. The dark horizontal line across each cluster shows the result of using the mean of all 100 random runs. The width of each cluster reflects the density of AUC among the 100 random runs at that bin. Bin sizes were of 0.02 AUC intervals. Note that there is a large drop in accuracy (AUC) from configuration VI to VII; this shows the difference between single- and multi-mineral classifications. See Table 2 for a description of each configuration that varies the modalities used as input into the model and the preprocessing done on Raman spectra.

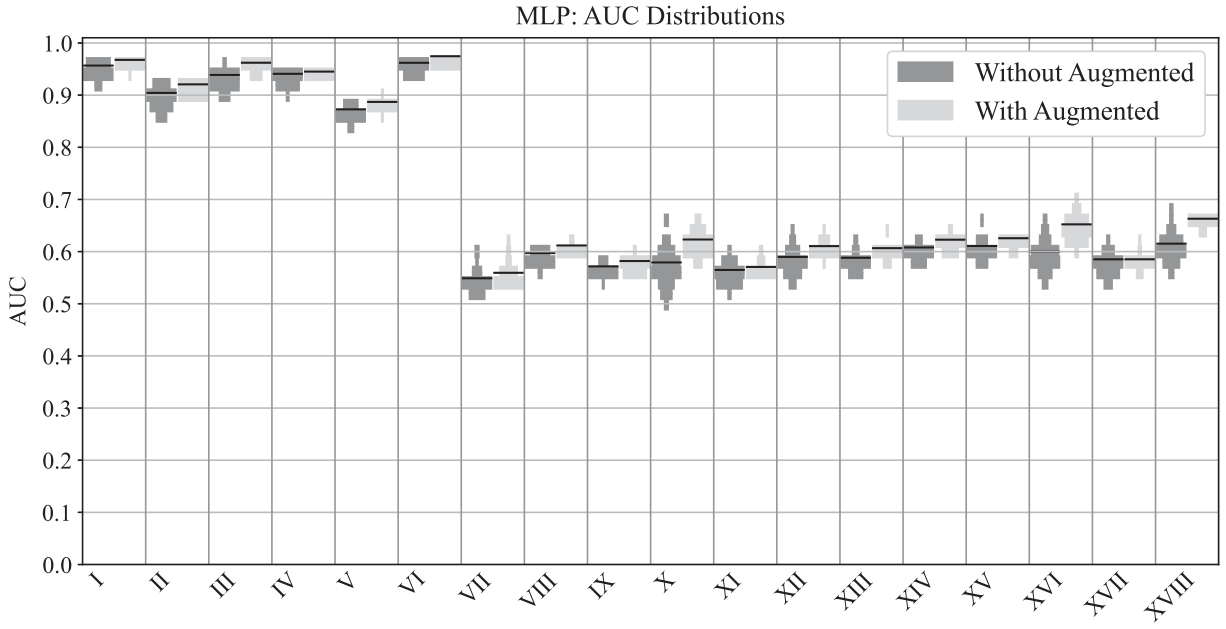


**Figure 5.** Results of each configuration using a Multi-Layer Perceptron (MLP) model. Not using augmented spectra in the training process corresponds to the left-sided distribution in each column and using augmented to the right. The dark horizontal line across each cluster shows the result of using the mean of all 100 random runs. The width of each cluster reflects the density of AUC among the 100 random runs at that bin. Bin sizes were of 0.02 AUC intervals. Note that there is a large drop in accuracy (AUC) from configuration VI to VII; this shows the difference between single- and multi-mineral classifications. See Table 2 for a description of each configuration that varies the modalities used as input into the neural network and the preprocessing done on Raman spectra.

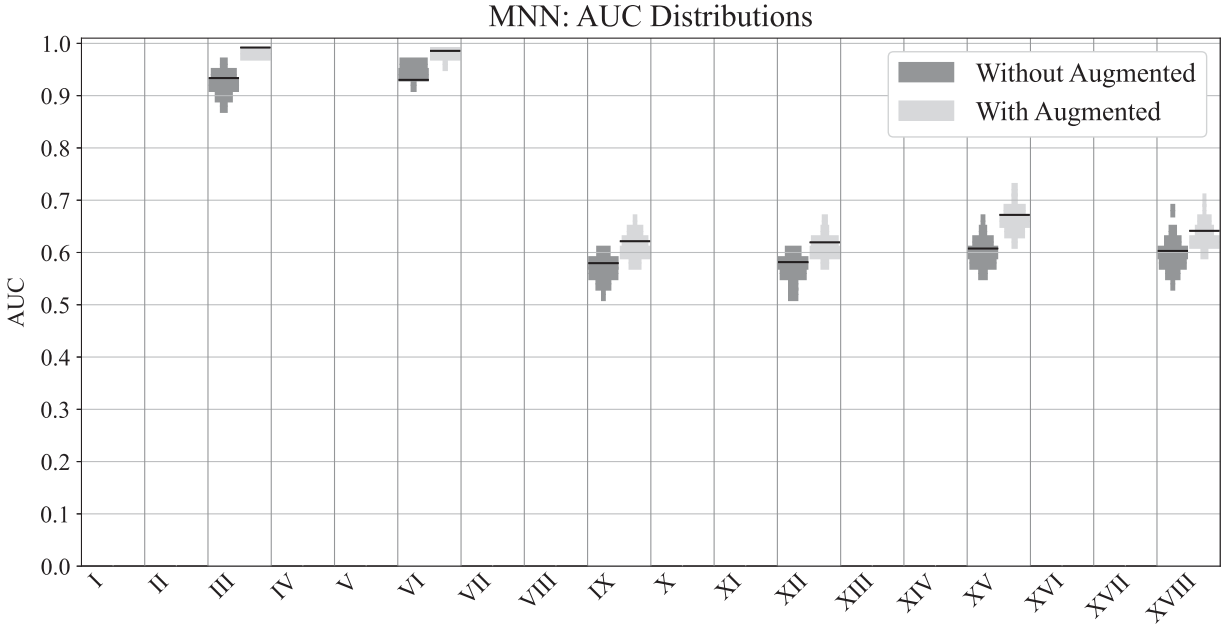


**Figure 6.** Results of each configuration using our Multimodal Neural Network (MNN) model. Not using augmented spectra in the training process corresponds to the left-sided distribution in each column and using augmented to the right. Note that the columns missing from this figure are those with single modality input. The dark horizontal line across each cluster shows the result of using the mean of all 100 random runs. The width of each cluster reflects the density of AUC among the 100 random runs at that bin. Bin sizes were of 0.02 AUC intervals. Note that there is a large drop in accuracy (AUC) for configurations after VI; this shows the difference between single- and multi-mineral classifications. See Table 2 for a description of each configuration that varies the modalities used as input into the neural network and the preprocessing done on Raman spectra.

**Table 3.** One-sided hypothesis test results

| Compared distributions | Single or multi | p-val from T-test $H_a$: $\overline{AUC}_1 > \overline{AUC}_2$ | $\overline{AUC}_1$ | $\overline{AUC}_2$ | $\sigma_1$ | $\sigma_2$ |
|---|---|---|---|---|---|---|
| augmented (1) | Single-mineral | 0.0 | 0.95 | 0.92 | 0.036 | 0.053 |
| not augmented (2) | Multi-mineral | 0.0 | 0.63 | 0.58 | 0.038 | 0.037 |
| with PCA (1) | Single-mineral | 1.0 | 0.93 | 0.94 | 0.051 | 0.045 |
| no PCA (2) | Multi-mineral | 1.0 | 0.60 | 0.61 | 0.045 | 0.042 |
| with mean (1) | Single-mineral | N/A | N/A | N/A | N/A | N/A |
| no mean (2) | Multi-mineral | 0.0 | 0.62 | 0.59 | 0.046 | 0.036 |
| mode A (1) | Single-mineral | 0.0 | 0.95 | 0.88 | 0.022 | 0.050 |
| mode B (2) | Multi-mineral | 0.0 | 0.60 | 0.59 | 0.045 | 0.039 |
| modes A and B (1) | Single-mineral | 0.0 | 0.96 | 0.95 | 0.023 | 0.022 |
| mode A (2) | Multi-mineral | 0.0 | 0.62 | 0.60 | 0.040 | 0.045 |
| modes A and B (1) | Single-mineral | 0.0 | 0.96 | 0.88 | 0.023 | 0.050 |
| mode B (2) | Multi-mineral | 0.0 | 0.62 | 0.59 | 0.040 | 0.039 |
| MLP (1) | Single-mineral | 0.0 | 0.94 | 0.92 | 0.033 | 0.060 |
| LR (2) | Multi-mineral | 0.83 | 0.60 | 0.60 | 0.030 | 0.056 |
| MNN (1) | Single-mineral | 0.0 | 0.97 | 0.92 | 0.017 | 0.060 |
| LR (2) | Multi-mineral | 0.0 | 0.63 | 0.60 | 0.029 | 0.056 |
| MNN (1) | Single-mineral | 0.0 | 0.97 | 0.94 | 0.017 | 0.035 |
| MLP (2) | Multi-mineral | 0.0 | 0.63 | 0.60 | 0.029 | 0.050 |

uses an MNN, dual-band excitation lasers, augmented spectra, and averaged all spectra from each sample (model configuration XV). The selected trained neural networks received test AUC scores of 0.997 and 0.733 for the single- and multi-mineral classifier, respectively.

The distributions shown in Figures 8 and 9 illustrate that the density of responses between ground truth positives vs. negatives change as the response value increases. Thus, Figures 8 and 9 show how different error metrics will change by sliding the cutoff value. Different regions can be identified for certain and uncertain classifications—where there is a high mix of ground truth positives and negatives. For example, in Figure 8, if the cutoff value for mica slid more to the right, then it will decrease the number of false positives but increase the number of false negatives. After selecting cutoff values for the selected neural networks, by maximizing the balanced accuracy in the absence of a defined cost function, we then made final classifications, which we can then evaluate using typical error metrics such as accuracy, recall, and precision.

Table 4 shows a confusion matrix resulting from leave-one-out cross-validation. These leave-one-out results reflect classification accuracy for co-registered dual-band Raman spectroscopy on pure mineral spectra collected from samples using the same spectrometer. Table 5 shows a confusion matrix with results from training on the SSI spectra and testing on the RRUFF spectra. These results reflect pure-mineral classification accuracy when using two different Raman excitation lasers on samples collected from a different spectrometer other than the training set. Figure 10 shows all the misclassified RRUFF spectra for both wavelengths. Table 6 shows various error metrics important for multi-classification problems. These results reflect multi-classification accuracy for co-registered dual-band Raman spectroscopy on rock spectra collected from samples using the same spectrometer.

## Discussion

In Figures 4–6, several clusters appear Gaussian, while others slightly deviate from this shape. However, most of these clusters have a single mode showing the expected AUC score, with vertical tails stretching to either side illustrating variance.

Hypothesis testing, as listed in Table 3, leads to several conclusions about which processes improve performance:

1. using augmented spectra in the training process results in a significantly higher AUC for both single and multi-mineral classification;

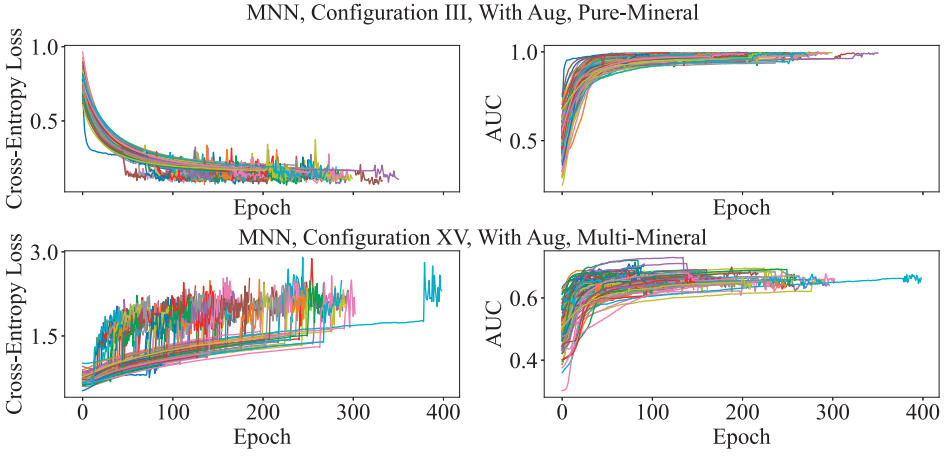


**Figure 7.** (**a**, top) Learning curves from 100 independent runs for the selected single-classifier configuration. (**b**, bottom) Learning curves from 100 independent runs for the selected multi-classifier configuration. Each learning curve shows accuracy as a function of training time. An ideal learning curve will smoothly converge to the global optimum.

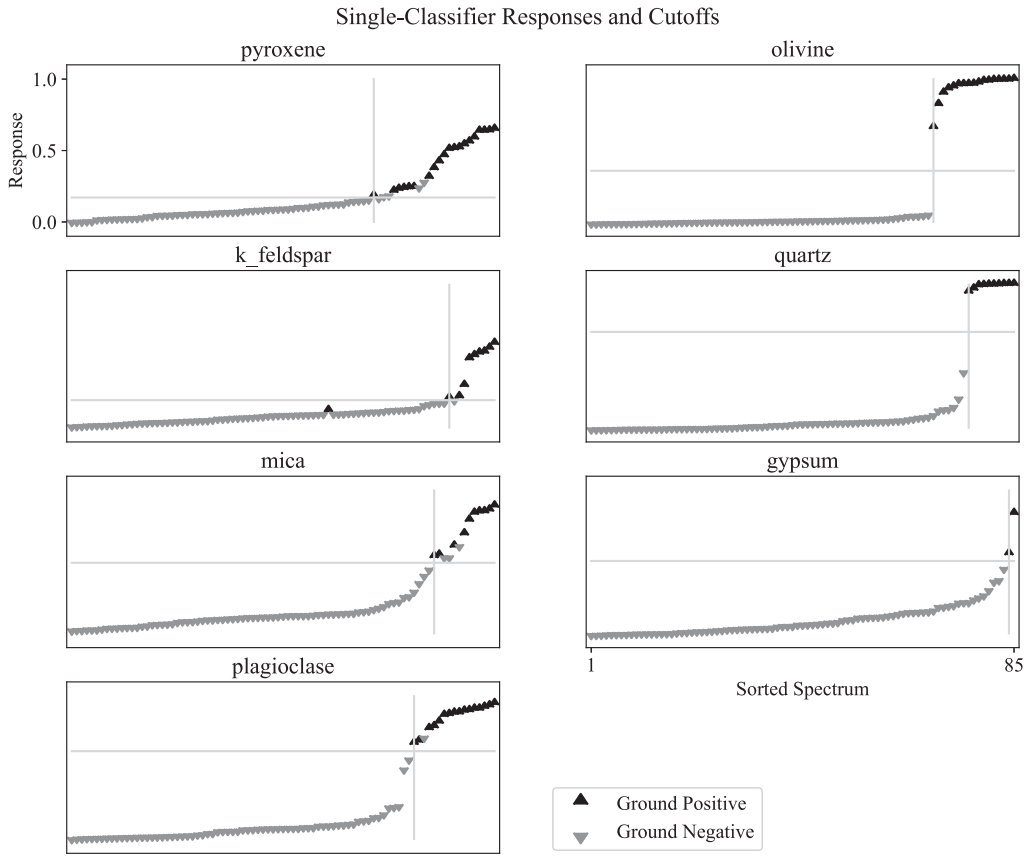


**Figure 8.** Model responses for the single-mineral classifier, sorted from lowest to highest. The cutoff values, indicated by intersecting lines, were found by optimizing balanced accuracy. Responses shown were from testing on RRUFF pure-mineral samples. This illustrates that as the cutoff value slides to the left or right, the number of false negatives and positives accordingly changes.

2. using PCA does not result in significantly higher AUC for either single- or multi-mineral classification;

3. using mean spectra results in a significantly higher AUC for multi-mineral classification;

4. using only the 532 nm spectra results in a significantly higher AUC than using only the 785 nm spectra for both single- and multi-mineral classification;

5. using both the 532 and 785 nm spectra results in a significantly higher AUC than using only the 532 nm spectra for both single- and multi-mineral classification;

6. using both the 532 and 785 nm spectra results in a significantly higher AUC than using only the 785 nm spectra for both single- and multi-mineral classification;

7. using an MLP results in a significantly higher AUC than using LR for single-mineral classification but does not show significant improvement for multi-mineral classification;

8. using our MNN results in a significantly higher AUC than using LR for both single- and multi-mineral classification;

9. using our MNN results in significantly higher AUC than using an MLP for both single- and multi-mineral classification.

All presented model configurations significantly improve over their predecessors, except for PCA. This is likely because the basis for PCA was calculated from the pure-mineral SSI spectra (training set) and then used as a change of basis for both the pure-mineral RRUFF and multi-mineral SSI spectra (testing sets). Using PCA may be beneficial if the classifier is allowed to calibrate the principal component basis on sampled data from the novel environment and the models are retrained in situ.

Figures 4–6. and Table 3 show low AUC averages for multi-mineral classification using the SSI rock samples, especially compared to single-mineral classification using the RRUFF pure minerals. This highlights the large gap in difficulty between these two types of classifications. The low AUC scores for multi-mineral classification, those near 0.5, are largely affected by the variations in model configurations that were being evaluated against each other. These are shown to illustrate the effectiveness and necessity of our methods. The global maximum values for an individual model in AUC between the single- and multi-mineral classifications are 0.997 and 0.733, respectively. These maximum values are more reflective of a final model.

Recall that a learning curve shows the model performance vs. training iteration. A surprising feature from the learning curves shown in Figure 7 is that, for the multi-mineral case, AUC can improve with increasing cross-entropy, the loss function used to train the neural network (see Online Materials[1] Appendix B). This is counter-intuitive but is geometrically explained as the distribution of inferences shifting down while simultaneously

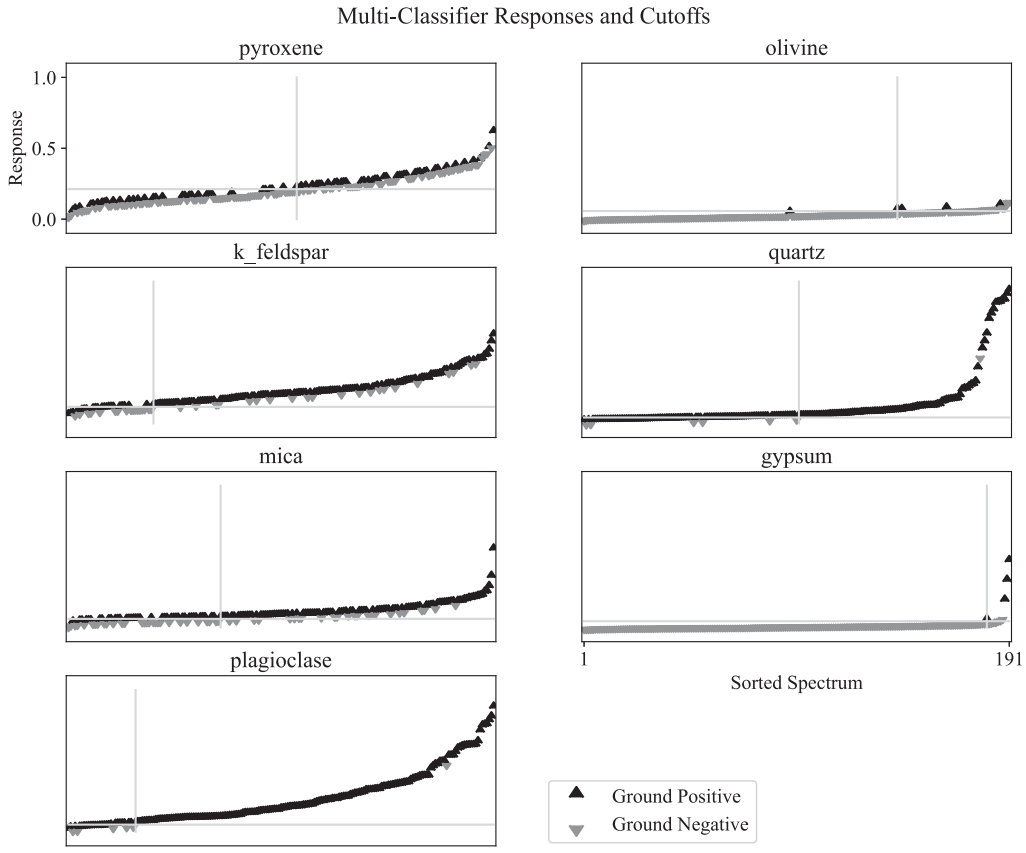


**Figure 9.** Model responses for the multi-mineral classifier, sorted from lowest to highest. The cutoff values, indicated by intersecting lines, were found by optimizing balanced accuracy. Responses shown were from testing on SSI multi-mineral samples. This illustrates that as the cutoff value slides to the left or right, the number of false negatives and positives accordingly changes.

**Table 4.** Single-mineral classifier leave-one-out results

| Mineral | Pyroxene | Olivine | K-feldspar | Quartz | Mica | Gypsum | Plagioclase | % |
|---|---|---|---|---|---|---|---|---|
| **Pyroxene** | 33 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| **Olivine** | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 100 |
| **K-feldspar** | 3 | 0 | 11 | 0 | 0 | 0 | 0 | 78.6 |
| **Quartz** | 0 | 0 | 0 | 18 | 0 | 0 | 0 | 100 |
| **Mica** | 1 | 0 | 0 | 0 | 10 | 0 | 0 | 90.9 |
| **Gypsum** | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 100 |
| **Plagioclase** | 0 | 0 | 0 | 0 | 0 | 0 | 47 | 100 |
| **%** | 89.2 | 100 | 100 | 100 | 100 | 100 | 100 | 96.5 |

**Table 5.** Single-mineral classifier test results

| Mineral | Pyroxene | Olivine | K-feldspar | Quartz | Mica | Gypsum | Plagioclase | % |
|---|---|---|---|---|---|---|---|---|
| **Pyroxene** | 15 | 0 | 0 | 0 | 5 | 0 | 0 | 75.0 |
| **Olivine** | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 100 |
| **K-feldspar** | 0 | 0 | 7 | 0 | 0 | 0 | 3 | 70.0 |
| **Quartz** | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 100 |
| **Mica** | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 100 |
| **Gypsum** | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 100 |
| **Plagioclase** | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 100 |
| **%** | 100 | 100 | 100 | 100 | 66.7 | 100 | 84.2 | 90.6 |

spreading out. The downward shift in responses is likely because the multi-mineral spectra used in the validation set do not closely resemble the pure-mineral spectra used in the training set. From the left panels in Figure 7, there are large fluctuations in cross-entropy that can arise after some number of epochs—as visible by the highly erratic regions of the learning curve. Such regions correspond to when the weights are unfrozen in the "hidden layer 1" of the MNN model (see Online Materials[1] Appendix B), and the location of these regions varies by exact epoch number due to convergence rates. In such regions, the effect on AUC highly varies, showing that these are regions of high uncertainty. For this reason, we opt not to unfreeze weights in the final models.

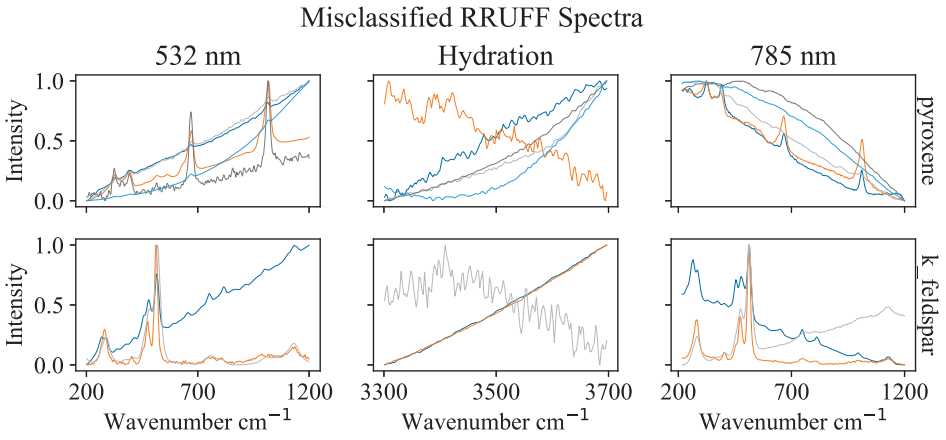


**Figure 10.** The 8 misclassified RRUFF spectra measured from pure-mineral samples. This group of 5 spectra measured from pyroxene minerals were misclassified as mica. This group of 3 spectra measured from potassium feldspar (K-feldspar) minerals were misclassified as plagioclase. There are anomalies visible in these spectra, such as unexpected background fluorescence patterns and otherwise lower signal-to-noise ratios than those observed in Figure 1.

**Table 6.** Multi-mineral classifier test results

| | Pyroxene | Olivine | K-feldspar | Quartz | Mica | Gypsum | Plagioclase | All |
|---|---|---|---|---|---|---|---|---|
| **AUC** | 0.53 | 0.77 | 0.63 | 0.66 | 0.66 | 0.99 | 0.79 | 0.72 |
| **True positive** | 40 | 4 | 132 | 94 | 99 | 4 | 160 | 533 |
| **False negative** | 32 | 1 | 21 | 90 | 39 | 0 | 25 | 208 |
| **False positive** | 49 | 47 | 21 | 1 | 24 | 7 | 1 | 150 |
| **True negative** | 70 | 139 | 17 | 6 | 29 | 180 | 5 | 446 |
| **Recall** | 0.56 | 0.80 | 0.86 | 0.51 | 0.73 | 1.0 | 0.87 | 0.72 |
| **Specificity** | 0.59 | 0.75 | 0.45 | 0.86 | 0.55 | 0.96 | 0.83 | 0.75 |
| **Precision** | 0.45 | 0.08 | 0.86 | 0.99 | 0.82 | 0.36 | 0.99 | 0.78 |
| **Sensitivity** | 0.69 | 0.99 | 0.45 | 0.06 | 0.42 | 1.0 | 0.17 | 0.68 |
| **Accuracy** | 0.58 | 0.75 | 0.78 | 0.52 | 0.67 | 0.96 | 0.86 | 0.73 |
| **$F_1$-score** | 0.50 | 0.14 | 0.86 | 0.67 | 0.76 | 0.53 | 0.93 | 0.75 |

When reviewing the results for single-mineral classification, Table 5 indicates that the only misclassified RRUFF spectra were a group of five pyroxene spectra classified as mica and a group of three potassium feldspar spectra classified as plagioclase. Pyroxene is a class of minerals with several diverse chemical compositions that create strongly varied responses—see the pyroxene spectra in Figure 1. The potassium feldspar spectra are likely misclassified due to the different patterns observed from the 785 nm laser between the RRUFF and SSI spectrometers—which can be seen when comparing the potassium feldspar spectra in Figure 10 (RRUFF) to those in Figure 1 (SSI).

The mineral gypsum is classified with nearly perfect accuracy for multi-mineral classification, while pyroxene struggles to be properly classified with an AUC near 0.5. This is expected behavior, as gypsum is typically used to calibrate Raman spectrometers due to their strong response, and pyroxene is problematic, as seen in the single-mineral classification results. Thus, our multimodal neural network methods are beneficial for identifying all mineral groups in rocks except for pyroxene. Generally, our models find the minerals easiest to predict (in decreasing order of AUC score) are gypsum, plagioclase, olivine, quartz, mica, potassium feldspar, and last pyroxene.

## Applications to the Moon and Mars

Our primary application of such methods is to increase the scientific return of instruments on planetary surface missions, especially to the Moon and Mars. Raman instruments will likely be deployed to the Moon on Artemis III (expected 2026 launch) and Artemis IV (expected 2028 launch) to be used by rovers and by suited astronauts to explore the lunar surface. Instead of having the astronauts try to figure out the minerals contained in a spectrum, the algorithms can return an identification of the minerals contained in the sample and the spectrum returned to users more experienced in interpreting spectra. Astronauts will likely encounter silicate minerals, such as plagioclase, pyroxene, and olivine, which make up >90% by volume of most lunar rocks (Wang et al. 1995; Papike et al. 1991). Other minerals also occur in lunar soils and rocks but are rare. These include quartz, cristobalite, tridymite, potassium feldspar, and zircon. However, oxide minerals, such as ilmenite, spinel, and armalcolite, are concentrated in some mare basalts and could be used as ores for resource extraction at lunar bases (Papike et al. 1991). In this paper, plagioclase, pyroxene, olivine, quartz, and potassium feldspar, all found on the Moon, were tested and identified by our automated spectral algorithm. Similar results would be expected for the surface exploration of Mars as well.

Our models require no more than 100 megabytes, and our data inputs (a dual-band Raman spectrum) are no larger than 16 kilobytes. Estimating the computational time it would take to make an inference is non-trivial, as there is no direct conversion between floating point operations and processor speeds. The BAE RAD750 processor installed on the Mars 2020 rover can perform over 200 million instructions per second. Appropriately, our MNN models take roughly $10^3$–$10^4$ floating point operations to make one classification. Thus, our machine learning model is lightweight and practical compared to existing planetary surface missions.

## Implications

In any geologic field exploration, the explorer, whether robot or human, must decide which rocks or soils are worthy of further attention and whether to include them in their sample collection. The Apollo astronauts made decisions based on gross shape or size of the rock or the color of the soil (for example, the famous orange soil of Apollo 17). However, in most lunar exploration cases, there is little visual feedback to guide the explorer from inside a spacesuit or from a rover camera. In addition, performing simple tests used by terrestrial field geologists, such as tasting a rock for halite, using a drop of dilute HCl to confirm a $CaCO_3$-rich rock, scratching the rock surface to test for hardness, or even scratching a fine-grained rock gently across their teeth to determine very fine sand and silt from clay grain sizes, would not be possible from the surface of the Moon or Mars. Therefore, the ability to quickly classify minerals, rocks, and sediment samples on a planetary surface could greatly enhance the science return from Mars rovers or suited astronauts on the Moon. The spectra returned by Raman spectrometers can be difficult to interpret in the field. Automated spectral algorithms provide quick opinions/results of the minerals contained in spectra, allowing the explorer to make informed decisions during their field reconnaissance. When such automated classifiers are integrated with Raman spectrometers, onboard classifications can improve decision making (e.g., where and how to probe next) without the delay in requiring multiple command cycles to interpret the data, thus improving science yield.

Integrating multiple sensors and fusing the resulting data can improve the robustness of autonomous mineral classification, as shown in this paper using two co-registered laser excitations in a single Raman spectrometer. However, extending these presented methods to integrate other data sources can further improve the science return. Each data source adds a unique perspective, aiding in the discovery of new correlations used to map provided input to desired output. The key advantage of using a neural network to learn such a mapping, as we do, is that high-performance computing is used during training, in which the algorithm optimizes model parameters with the high dimensional feature-space while keeping the number of computations low during inference, as to make the algorithms deployable to the resource-constrained devices typical in remote surveyors. We posit that integrating larger databases with multiple sensors will yield robust IDU systems capable of autonomously exploring remote planetary surfaces.

## Acknowledgments and Funding

We thank Patrick Freeman and Paige Morkner for obtaining and analyzing Raman spectra and Jason Angell for coding scripts to plot and obtain Raman spectra. We thank Job Bello for developing and supplying our dual excitation (535 and 785 nm) Raman spectrometer. We thank Shawn Hart, David Wettergreen, Nancy Hinman, and Ted Vassalo for collecting and sharing samples from their field sites. DJ Minerals supplied several of the mineral and rock samples for this project. We thank Shawn Hart, Paige Morkner, Alicia Horton, Khanh Luu, and Richard Nelson for helping to analyze hand samples. We thank all the contributors to the RRUFF database. We thank Sascha Ishikawa for developing the initial mineral classifier and early comments for improvement. Funding was provided by the NASA Ames Internship program, the NASA Astrobiology Program Grant #NNX15BB01, and the National Science Foundation grant DUE-1930546.

## References Cited

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M. and Ghemawat, S. (2015) TensorFlow: Large-scale machine learning on heterogeneous distributed systems. https://doi.org/10.48550/arXiv.1603.04467.

Ambruster, T. and Danisi, R.M. (2015) The power of databases: The RRUFF project. In T. Armbruster and R.M. Danisi, Eds., Highlights in Mineralogical Crystallography, p. 1–30. De Gruyter.

Beegle, L., Bhartia, R., White, M., DeFlores, L., Abbey, W., Wu, Y.H., Cameron, B., Moore, J., Fries, M., Burton, A. and Edgett, K.S. (2015) SHERLOC: Scanning habitable environments with Raman and luminescence for organics and chemicals. 2015 IEEE Aerospace Conference, p. 1–11.

Bengio, Y., Laufer, E., Alain, G. and Yosinski, J. (2014) Deep generative stochastic networks trainable by backprop. Proceedings of the 31st International Conference on Machine Learning, 32, 226–234.

Bergstra, J., Yamins, D., and Cox, D.D. (2013) Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. Proceedings of the 30th International Conference on Machine Learning, p. 115–123.

Berlanga, G, Williams, Q., and Temiquel, N. (2022) Convolutional neural networks as a tool for Raman spectral mineral classification under low signal, dusty mars conditions. Earth and Space Science, 9, 10.1029/2021EA002125.

Chollet, F. (2015) Keras. Software available from keras.io.

Clevert, D., Unterthiner, T., and Hochreiter, S. (2015) Fast and accurate deep network learning by exponential linear units (ELUs).

Cochrane, C.J. and Blacksberg, J. (2015) A fast classification scheme in Raman spectroscopy for the identification of mineral mixtures using a large database with correlated predictors. *IEEE Transactions on Geoscience and Remote Sensing,* 53, 4259–4274, https://doi.org/10.1109/TGRS.2015.2394377.

Crow, P., Barrass, B., Kendall, C., Hart-Prieto, M., Wright, M., Persad, R., and Stone, N. (2005) The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines. *British Journal of Cancer,* 92, 2166–2170, https://doi.org/10.1038/sj.bjc.6602638.

Downs, R.T. (2006) The RRUFF project: An integrated study of the chemistry, crystallography, Raman and infrared spectroscopy of minerals. Program & Abstracts: 19th General Meeting of the International Mineralogical Association, Kobe, Japan, July 23–28, 2006. IMA.

Dyar, M.D., Breitenfeld, L.B., Carey, C.J., Bartholomew, P., Tague, T.J. Jr., Wang, P., Mertzmann, S., Byrne, S.A., Crowley, M.C., Leight, C., and others. (2016) Interlaboratory and cross-instrument comparison of Raman spectra of 96 minerals. *Lunar and Planetary Science,* 46, 2240.

Francis, R., Estlin, T., Doran, G., Johnstone, S., Gaines, D., Verma, V., Burl, M., Frydenvang, J., Montaño, S., Wiens, R.C. and Schaffer, S. (2017) AEGIS autonomous targeting for ChemCam on Mars Science Laboratory: deployment and results of initial science team use. *Science Robotics,* 2.7, eaan4582.

Freeman, J.J., Wang, A., Kuebler, K.E., Jolliff, B.L., and Haskin, L.A. (2008) Characterization of natural feldspars by Raman spectroscopy for future planetary exploration. *Canadian Mineralogist,* 46, 1477–1500, https://doi.org/10.3749/canmin.46.6.1477.

Glorot, X. and Bengio, Y. (2010) Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the international conference on artificial intelligence and statistics, Journal of Machine Learning Research, 9, 249–256.

He, K., Zhang, X., Ren, S. and Sun, J. (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, 1026–1034.

Howley, T., Madden, M.G., O'Connell, M.L. and Ryder, A.G. (2005) The effect of principal component analysis on machine learning accuracy with high dimensional spectral data. In A. Macintosh, R. Ellis, and T. Allen, Eds., Applications and Innovations in Intelligent Systems XIII. SGAI 2005. Springer.

Ishikawa, S.T. and Gulick, V.C. (2013) An automated mineral classifier using Raman spectra. *Computers & Geosciences,* 54, 259–268, https://doi.org/10.1016/j.cageo.2013.01.011.

Johnsen, T.K., Marley M.S., and Gulick V.C. (2020) A multilayer perceptron for obtaining quick parameter estimations of cool exoplanets from geometric albedo spectra. *Publications of the Astronomical Society of the Pacific,* 132, 044502.

Kingma, D.P. and Ba, J. (2014) Adam: A method for stochastic optimization. Conference of Learning Representations.

Lewis, I.R. and Edwards, H. (2001) Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line, 1072 p. CRC Press.

Liu, Y., Wang, A., and Freemen, J.J. (2009) Raman, MIR, and NIR spectroscopic study of calcium sulfates: gypsum, bassanite, and anhydrite. 40th Lunar and Planetary Science Conference, (Lunar and Planetary Science XL), March 23–27, 2009, The Woodlands, Texas, 2128.

Lopez-Reyes, G., Sobron, P., Lefebvre, C. and Rull, F. (2014) Multivariate analysis of Raman spectra for the identification of sulfates: Implications for ExoMars. *American Mineralogist,* 99, 1570–1579, https://doi.org/10.2138/am.2014.4724.

Maurice, S., Wiens, R.C., Bernardi, P., Caïs, P., Robinson, S., Nelson, T., Gasnault, O., Reess, J.M., Deleuze, M., Rull, F., and others. (2021) The SuperCam instrument suite on the Mars 2020 rover: Science objectives and mast-unit description. *Space Science Reviews,* 217, 47, https://doi.org/10.1007/s11214-021-00807-w.

Nair, V. and Hinton, G.E. (2010) Rectified linear units improve restricted Boltzmann machines. Proceedings of the 27th International Conference on Machine Learning, p. 807–814. International Conference on Machine Learning-10.

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H. and Ng, A.Y. (2011) Multimodal deep learning. Proceedings of the 28th International Conference on Machine Learning, p. 689–696. International Conference on Machine Learning-11.

Papike, J., Taylor, L., and Simon, S. (1991) Lunar Sourcebook: A Users Guide to the Moon, p. 121–182. Cambridge University Press.

Prechelt, L. (1998) Early stopping - But when? In G. Montavon, G.B. Orr, and K-R. Müller, Eds., Neural Networks: Tricks of the Trade, 55–69. Springer.

Rull, F., Maurice, S., Hutchinson, I., Moral, A., Perez, C., Diaz, C., Colombo, M., Belenguer, T., Lopez-Reyes, G., Sansano, A. and Forni, O. (2017) The Raman laser spectrometer for the ExoMars rover mission to Mars. *Astrobiology,* 17, 627–654.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986) Learning representations by back-propagating errors. *Nature,* 323, 533–536, https://doi.org/10.1038/323533a0.

Ryder, A.G. (2002) Classification of narcotics in solid mixtures using principal component analysis and Raman spectroscopy. *Journal of Forensic Sciences,* 47, 275–284, https://doi.org/10.1520/JFS15244J.

Salakhutdinov, R., Mnih, A., and Hinton, G. (2007) Restricted Boltzmann machines for collaborative filtering. Proceedings of the 24th International Conference on Machine Learning, p. 791–798.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014) Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research,* 15, 1929–1958.

Verma, V., Maimone, M.W., Gaines, D.M., Francis, R., Estlin, T.A., Kuhn, S.R., Rabideau, G.R., Chien, S.A., McHenry, M.M., Graser, E.J. and Rankin, A.L. (2023) Autonomous robotics is driving Perseverance rover's progress on Mars. Science Robotics, 8, eadi3099.

Vincent, P., Larochelle, H., Bengio, Y. and Manzagol, P.A. (2008) Extracting and composing robust features with denoising autoencoders. Proceedings of the 25th International Conference on Machine Learning, p. 1096–1103.

Wang, A., Jolliff, B.L., and Haskin, L.A. (1995) Raman spectroscopy as a method for mineral identification on lunar robotic exploration missions. *Journal of Geophysical Research: Solid Earth,* 100 (E10), 21189–21199, https://doi.org/10.1029/95JE02133.

Wang, A., Freeman, J.J., Jolliff, B.L., and Chou, I.M. (2006) Sulfates on Mars: A systematic Raman spectroscopic study of hydration states of magnesium sulfates. *Geochimica et Cosmochimica Acta,* 70, 6118–6135, https://doi.org/10.1016/j.gca.2006.05.022.

Wiens, R.C., Maurice, S., Robinson, S.H., Nelson, A.E., Cais, P., Bernardi, P., Newell, R.T., Clegg, S., Sharma, S.K., Storms, S., and others. (2021) The SuperCam instrument suite on the NASA Mars 2020 rover: Body unit and combined system tests. *Space Science Reviews,* 217, 4, https://doi.org/10.1007/s11214-020-00777-5.

Xie, C., Mace, J., Dinno, M.A., Li, Y.Q., Tang, W., Newton, R.J., and Gemperline, P.J. (2005) Identification of single bacterial cells in aqueous solution using confocal laser tweezers Raman spectroscopy. *Analytical Chemistry,* 77, 4390–4397, https://doi.org/10.1021/ac0504971.

Manuscript received May 23, 2023
Manuscript accepted July 5, 2024
Accepted manuscript online July 16, 2024
Manuscript Handled by Janice Bishop

## Endnotes:

[1]Deposit item AM-25-59072. Online Materials are free to all readers. Go online, via the table of contents or article view, and find the tab or link for supplemental materials.