Imperial College London

Department of Earth Science and Engineering

MSc in Environmental Data Science and Machine Learning

Independent Research Project

Final Report

Hyperspectral Image Processing on Mars

by

Jie Wang

Email: jie.wang22@imperial.ac.uk

GitHub username: edsml-jw3222

Repository: https://github.com/ese-msc-2022/irp-jw3222.git

Supervisors:

Dr. Cedric John

Mr. Robert Platt

September 2023

Abstract

The quest for understanding Mars' surface necessitates high-quality hyperspectral imaging data. In such data, noise reduction is seen as an essential factor for achieving accurate insights. In this research, a methodology is introduced to enhance the clarity of Mars hyperspectral images. In the first part, a Logistic Regression model is incorporated for the accurate identification of bland pixels. Once identified, these bland pixels are employed in a column-wise pixel ratioing algorithm, devised to highlight features in the spectral data. Leveraging the inherent automation capabilities of machine learning, the presented approach demonstrates higher efficiency compared to traditional manual selection. Moreover, it provides more reliable outcomes than the median ratioing. This methodology offers a flexible and streamlined tool for the interpretation of Martian surface.

1. Introduction

The rich geological history and unique climate have positioned Mars as a significant target for space exploration. To deepen the understanding of the planet, the Mars Reconnaissance Orbiter (MRO) project was launched by NASA. The primary mission of the orbiter was to provide detailed maps for future exploration tasks and to study the climate and geology of Mars comprehensively (Zurek et al., 2007). Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) is a hyperspectral imager deployed on MRO, designed to capture the reflected light from the Martian surface. Unlike traditional imagers which detect only a limited range of colors, CRISM was developed to distinguish hundreds of wavelengths, spanning from ultraviolet to near-infrared (Murchie et al., 2007). This capability allows for a detailed understanding of the reflection and absorption features of materials. Since the MRO entered the orbit of Mars in 2006, extensive spectral data from CRISM has enriched Martian mineralogy and geological research.

To analyze these detailed data more effectively, summary parameters were introduced (Viviano-Beck et al., 2014). These parameters are designed to capture essential spectral features, such as the position, width, or depth of absorption bands, offering a simplified and quantified approach for describing mineral spectral features. By using this method, key information can be efficiently extracted from spectra data, aiding in mineral identification.

While summary parameters offer an efficient approach for analyzing spectral data, their reliability is hinged on the quality of the data. Noise can be introduced into the spectral features due to environment of Mars as well as CRISM itself. In Figure 1, a clean spectrum is compared with a noisy one. Addressing these noise factors is considered vital for the accurate identification of minerals. Manual denoising, median ratioing, and other methods are commonly employed. In this research, the potential of machine learning for addressing spectral noise is explored, and its performance is compared with these traditional methods.

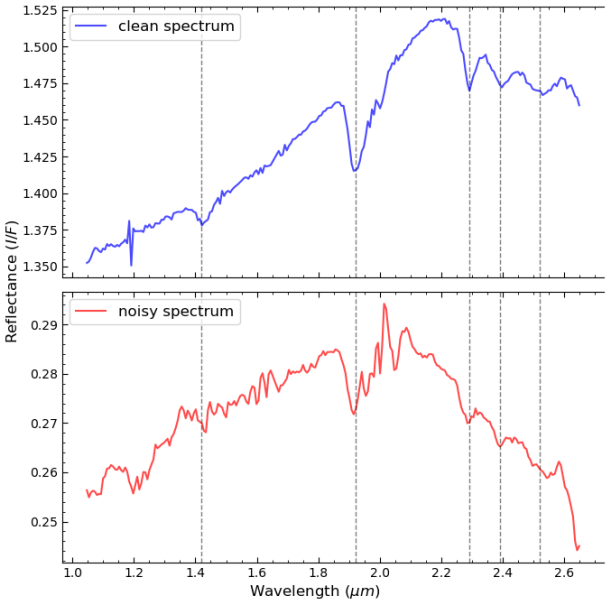

Figure 1. Clean and noisy spectra of Fe smectite. The vertical dashed lines indicate positions of its spectral features are located. The two features between wavelengths 1.4 μm and 2.0 μm are masked by noise, while the other features are weakened.

2. Background

Spectral data inevitably contain noise from various sources. This section we will discuss the environmental and instrumental sources, along with prevalent denoising techniques such as manual selection and median ratioing.

2.1 Sources and Properties of Spectral Noise

The environmental noise comes from the diverse environmental conditions on Mars's surface. Each pixel, representing a unique geographical region, may contain non-mineral elements such as dust, meteorological conditions (Bultel et al., 2015) that respond to infrared radiation at specific wavelengths, potentially obscuring the true spectral features of minerals or introducing irrelevant spectral features. As a result, each pixel's environmental noise is distinct, making it challenging to find a universal denoising method suitable for Mars' diverse conditions.

The CRISM instrument employs a push-broom imaging technique, capturing Mars's reflected light column-wise to create spectral images (Murchie et al., 2007). This continuous data collection method covers vast areas quickly but also result in 'spectral smile', which is caused by non-homogeneous spectral response along the across-track position (Ceamanos et al., 2009). This instrumental noise remains consistent within a single column but varies between columns, necessitating column-wise corrections to maintain data accuracy.

2.2 Reducing Spectral Noise

From a mathematical perspective, environmental noise consists of additive part and multiplicative part (Plebani et al., 2022). Additive part arises from random or systematic disturbances, added directly to the original spectrum. Given its random nature, its average value over extensive data approaches zero. One might assume that averaging spectral data would mitigate this type of noise, buy it can also average out any mineral absorption features, which are crucial for analysis. Therefore, additive noise is still not fully addressed in CRISM data. Given this challenge, this research primarily focuses on the multiplicative part.

Multiplicative noise arises from the reflective and absorptive features of non-minerals on the Martian surface. It modifies the original spectrum, amplifying in certain wavelength ranges while dampening in others. To reduce this noise, bland pixels are frequently employed as reference pixels. They are spectrally neutral and

lack mineral features but have similar environmental noise with target pixels. By ratioing the spectrum of target pixel with bland pixel, one can reduce multiplicative noise and enhance the true spectral features of the minerals (Bultel et al., 2015).

2.3 Traditional Methods and Limitations

Pixel ratioing depends on identifying suitable bland pixel. In traditional methods, summary parameters guide this manual selection. Pixels not emphasized by any summary parameter typically serve as candidates for the denominator, being considered bland and without distinct mineral features (Plebani et al., 2022). However, it's time-consuming when dealing with large datasets. Moreover, there's a risk of erroneously identifying pixels containing unknown minerals, which are not yet incorporated into summary parameters, as bland pixels. In this case, pixel ratioing can either mask these unknown mineral features from the target pixel or introduce misleading features if the target lacks this unknown mineral.

Beyond the traditional use of bland pixels for ratioing, an alternative approach is median ratioing. This method offers an automated solution to the challenge of noise reduction. Instead of using a bland pixel as the denominator for ratioing, median ratioing computes the median spectrum of the column containing the target pixel and uses it as the denominator. A notable drawback of this is that the median spectrum isn't necessarily spectrally neutral (Bultel et al., 2015) and thus has the risk of masking valuable spectral features. Therefore, while median ratioing provides a quick deployment advantage, its lack of discernment in choosing the denominator may not match the precision of manual selection.

Machine learning offers a promising way to address these challenges. When properly trained, these models can recognize complex patterns in the data that might be missed by traditional methods. Furthermore, machine learning's automated nature allows it to quickly process large amounts of data. The adjustable prediction threshold in machine learning models provides extra flexibility, allowing adjustments of bland level on different situations, which can deliver better accuracy than median ratioing.

3. Methods

This research introduces a three-phase method to reduce noise, as illustrated in Figure 2. First, selecting and processing appropriate data; second, classifying bland pixels through trained machine learning model; and lastly, pixel ratioing.

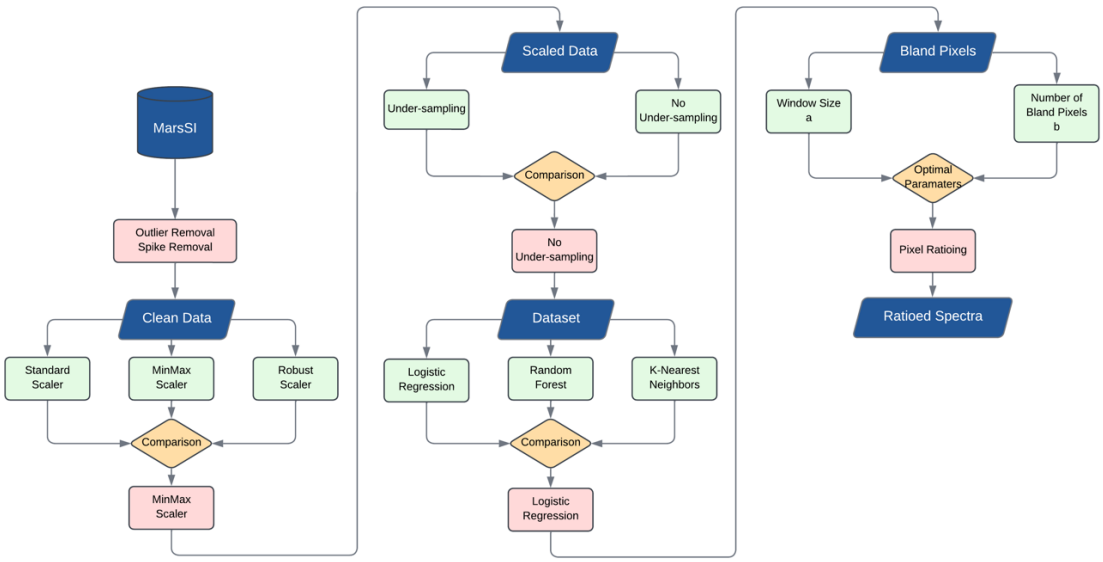

Figure 2. Flowchart illustrating the comprehensive research methodology and three-stage noise reduction process for Mars hyperspectral images.

3.1 Data Structure

CRISM offers a range of calibrated data products. To broaden the applicability of this research, we chose the widely used TRDR data product, and focuses on all 438 bands ranging from 1.0 to 4.0 micrometers. The data for this research was sourced from the MarsSI (Mars System of Information) database (Quantin-Nataf et al., 2018). The detailed dataset structure used for training can be found in Table 1.

Table 1
Structure of the dataset used for training.
Stored as a pandas dataframe, with dimensions of 909,499 × 442.

Name	Size	Description
image ids	337	Image ID associated with the pixel
labels	909,499	Category/Class of the pixel
coordinates	909,499 × 2	(x, y) coordinates of the pixel within the image
spectra	909,499 × 438	Spectral data of the pixel

Table 1. the image IDs, labels, and coordinates are sourced from Plebani et al. (2022).

3.2 Data Cleaning

Upon examination of the dataset, it was observed that the majority of values are concentrated around the 0-1 range, with some values slightly deviating beyond this interval. The concentration in this range is due to the use of the I/F unit, which ideally represents the ratio of the observed radiance to the calculated solar flux (Murchie et al., 2007), confining values within the 0-1 range under ideal conditions. However, in practice, factors such as instrument noise, atmospheric interference, or processing anomalies can cause minor deviations. Instances of constant values at 65535 were also detected, indicative of an invalid marker in CRISM data. To preserve the integrity of our dataset, a series of cleaning steps were initiated.

Firstly, values that significantly deviated from the theoretical 0 to 1 range were replaced with the average value of the image. Understanding that measurement precision and external interferences might introduce slight deviations, a tolerance was applied, extending the acceptable range to -1 to 2.

Next, to address abnormal spikes, the median filtering method from the SciPy library was used. With a window size of three and a threshold of five standard deviations for peak identification, any values surpassing this threshold were deemed anomalies and were substituted with the window's median. Importantly, considering the column-related instrumental noise discussed in section 2.1, this filtering was applied column-wise.

3.3 Bland Pixel Identification

The identification of bland pixels was pursued using machine learning models. In this approach, all other mineral categories were grouped into a single category, transforming the problem into a binary classification task. Three distinct models were chosen for evaluation: Logistic regression, which served as a baseline to probe the linear attributes of the data; Random Forest, known for its ability to capture nonlinear relationships; and KNN, which is often favored for handling high-dimensional data.

Considering the class distribution of the dataset, it comprised 586,576 mineral pixels (labeled as 0) and 322,877 bland pixels (labeled as 1). Given the evident class imbalance, the F1 score was primarily emphasized in the evaluation, with accuracy and recall also taken into account. The potential impact of class balance on model performance was examined by applying random under-sampling.

Various data scaling techniques, including the Standard, Min-Max, and Robust Scalers, were assessed for their influence on the performance of each model. To ensure an accurate measure of the model's generalization capability, a comparison was made between grouped cross-validation, where images served as distinct groups, and traditional cross-validation.

3.4 Pixel Ratioing

Considering the possible column-related instrumental noise from CRISM, our algorithm selects bland pixels from the same column as the target pixel. Specifically, the algorithm defines a window size parameter a, covering a pixels above and below the target. Within this window, the top b bland pixels are chosen by the machine learning model based on their likelihood, and their average spectrum is utilized as the denominator of ratioing.

The fine-tuning of parameters is recognized as essential for achieving optimal results. Window size establishes the search boundary for bland pixels. If it is set excessively large, pixels too distant from the target might be included, potentially leading to a misrepresentation of environmental noise. On the contrary, a value that is too small might not encompass enough bland pixels. Number of bland pixels, meanwhile, dictates the number of bland pixels to be utilized for ratioing. A high value might inadvertently encompass pixels with mineral features, while a low value might not adequately represent the environmental noise surrounding the target pixel.

For targets identified as specific regions rather than individual pixels, relevant parameters are also provided by our algorithm. Ratioing is applied to each pixel within the designated area, after which the averaged spectrum of these ratioed pixels is computed, yielding unified spectral features for the entire region.

4. Results

4.1 Machine Learning Model Training and Evaluation

To assess the generalization capability of the models, two cross-validation methods were employed: standard KFold and GroupKFold, both with five splits. While the standard KFold divides the dataset into five parts, the GroupKFold ensures that no image data is shared between the training and validation sets, preventing potential data leakage.

In Figure 3, distinct differences between the two validation methods are observed. Although similar training trends are exhibited by both, model performance appears to be overestimated by KFold. Notably, F1 scores approaching 1 were observed for Random Forest and KNN, while Logistic Regression achieved scores exceeding 0.9. In contrast, overfitting was indicated by GroupKFold in the Random Forest and KNN models with F1 scores hovering around 0.65. However, an F1 score of nearly 0.84 was achieved by Logistic Regression. Therefore, we chose GroupKFold for all subsequent performance evaluations.

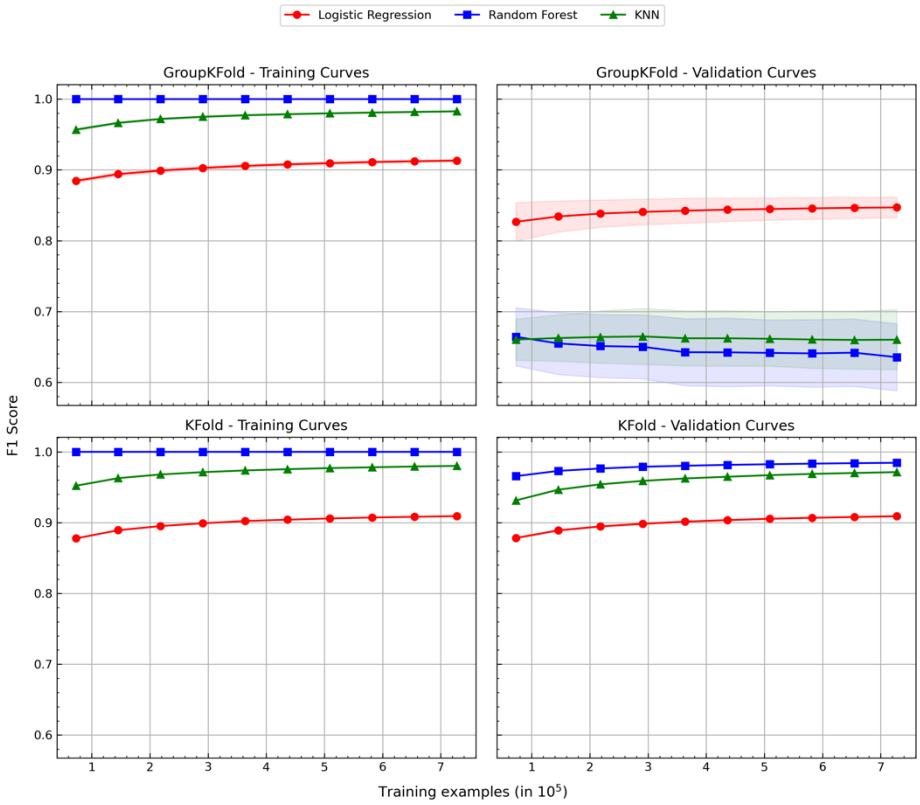

Figure 3. Training and validation curves of the three models: the top row presents results using GroupKFold, while the bottom row depicts outcomes with standard KFold.

The influence of various data scaling methods on model performance was further investigated. Specifically, the impacts of the Standard, Min-Max, and Robust Scaler on performance were evaluated and detailed in Table 2. Only minor performance fluctuations were observed across all three models when using different scalers. The Min-Max Scaler was chosen as the definitive scaling approach for this research, especially considering that some of the valid data slightly deviated from the theoretical 0-1 range, as mentioned in section 3.2. Utilizing the Min-Max Scaler allowed for the normalization of all values within the theoretical range.

Table 2
Comparison of F1 scores for different scaling methods.

Scaler	Logistic Regression	Random Forest	KNN
Standard Scaler	0.8560	0.6625	0.6658
Min-Max Scaler	0.8484	0.6597	0.6778
Robust Scaler	0.8560	0.6594	0.6528

Concerning the data imbalance issue, experiments were conducted using random under-sampling. To ensure a fair comparison, an independent test set was utilized to evaluate the models trained on imbalanced data against those trained on balanced data. All evaluation metrics are presented in Table 3. Decreased performance for both classes was consistently observed across all models with the use of under-sampling, suggesting that balancing the data may not necessarily enhance the model's capability in identifying a particular category. Given these findings, it was concluded that preserving the original data distribution might be more suited for this research.

Table 3
Comparison of F1 scores for balanced and unbalanced data.
Test set with 68 images, which has 65,584 bland pixels and 103,676 mineral pixels.

(a) Performance on Class 0: Mineral Pixel

Data	Logistic Regression	Random Forest	KNN
Balanced	0.91	0.51	0.65
Unbalanced	0.93	0.52	0.68

(b) Performance on Class 1: Bland Pixel

Data	Logistic Regression	Random Forest	KNN
Balanced	0.87	0.65	0.67
Unbalanced	0.89	0.65	0.68

From the analyses conducted, it was decided to retain the imbalanced data distribution, and the Min-Max Scaler was employed. Based on our analyses, we opted to retain the imbalanced data distribution and applied the Min-Max Scaler. Under these preset conditions, we fine-tuned the hyperparameters of the three models using the same test set mentioned above, and F1 score was chosen as the evaluation metric. All attempted hyperparameter combinations and their corresponding evaluation results are listed in Table 4. It's evident that the various hyperparameter combinations attempted did not significantly influence any model's performance.

Table 4
Comparison of F1 scores for Hyperparameter Tuning.

(a) Logistic Regression

Hyperparameter	saga	liblinear	newton-cg
C=1	0.90	0.91	0.91
C=10	0.90	0.91	0.91

(b) Random Forest

n estimators	50	100	200
F1 Score	0.60	0.59	0.59

(c) KNN

n neighbors	5	8	11
F1 Score	0.68	0.69	0.69

Ultimately, amongst the evaluated models, logistic regression was chosen for its superior performance. For the logistic regression model, the newton-cg solver was selected, while other parameters were kept at their default values. Table 5 showcases its average classification report of cross-validation over the entire dataset.

Table 5
Classification report for Logistic Regression.

	Precision	Recall	F1-score	Support
Class 0: Mineral	0.94	0.88	0.91	586576
Class 1: Bland	0.81	0.89	0.85	322877
Macro avg	0.87	0.89	0.88	909453
Weighted avg	0.89	0.89	0.89	909453

4.2 Evaluation of Pixel Ratioing

Four pixels were selected from four distinct images, each representing a different mineral, for pixel ratioing. The logistic regression model described in Section 4.1 was applied. To prevent data leakage, all pixels from these four images were excluded from the training set and the model was retrained. In addition to this, a comparison with median ratioing was also conducted. Figure 4 illustrates the comparative results. It's observed that the ratioing effect is pronounced for Fe-smectite, Kaolinite, and Jarosite. It enhances features in the ratioed spectra that might be subtle otherwise. However, for Alunite, the feature at the wavelength of 1.94 micrometers was overly amplified. On the other hand, median ratioing often inadvertently erases spectral features, with the ratioed spectrum shape not closely aligning with the reference spectrum. In the case of Alunite, median ratioing almost obliterated most of the features. It is evident that while using bland pixels for ratioing effectively reduced noise and enhanced features, the median spectrum tends to erase many useful features.

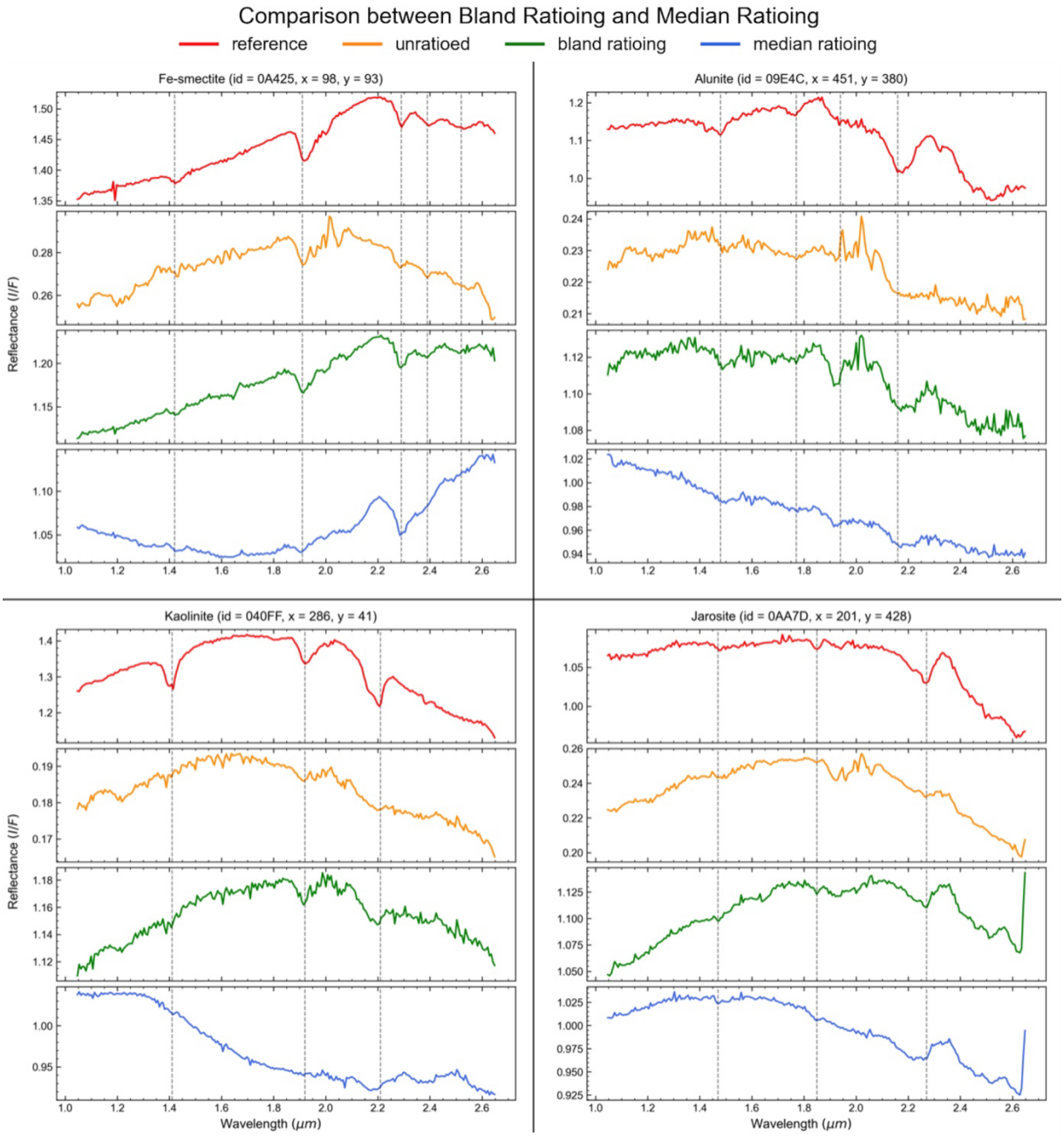

Figure 4. The dashed lines indicate the spectral feature positions of the corresponding minerals. Fe-smectite and Kaolinite used window size 50 and 3 bland pixels, while Alunite and Jarosite used window size 50 and 5 bland pixels. Each subplot has four curves, from top to bottom, the reference spectrum from the MRO CRISM Type Spectra Library, the original spectrum of the target pixel, the spectrum ratioed by bland pixel, and the spectrum ratioed by median spectrum.

5. Discussions

5.1 Comparison

A crucial part is the accurate identification of bland pixels. The Hierarchical Bayesian model (HBM), which was proposed by Plebani et al., was employed for a comparative analysis. This model was contrasted against our method using the test set derived from section 4.1, encompassing 65,584 bland pixels. It's essential to understand that the HBM functions differently from traditional classification models. Instead of directly classifying pixels as bland or mineral, each pixel is assigned a bland score by the HBM. To comparison, pixels were sorted based on these scores in a descending order. The top 65,584 pixels were then labeled as bland, while all the subsequent pixels were labeled as mineral. The outcome of this comparison is presented in Table 6. It was observed that the performance rendered by the HBM was not ideal.

Table 6
Comparison of Logistic Regression and Hierarchical Bayesian.
Test set with 68 images, which has 65,584 bland pixels and 103,676 mineral pixels.

(a) Performance of Logistic Regression

Class	Precision	Recall	F1-score
0: Mineral	0.93	0.92	0.93
1: Bland	0.88	0.89	0.89

(b) Performance of Hierarchical Bayesian

Class	Precision	Recall	F1-score
0: Mineral	0.51	0.51	0.51
1: Bland	0.22	0.22	0.22

5.2 Analysis

The results from our experiments on cross-validation and under-sampling have shown several interesting phenomena. In Figure 4, apart from the learning curves, variances were also plotted. Notably, a significant variance was observed in the validation curves of GroupKFold, whereas standard KFold didn't exhibit such a trend. This highlights a unique aspect of our training data. Of the 337 images utilized in our research, 260 contain only label 1, 12 have only label 0, and a mere 65 images feature both labels. The GroupKFold may split different types of images into training or validation sets, leading to substantial class distribution variations across the folds. This could be the reason behind the oscillations in the learning curve. On the other hand, the traditional KFold, due to data leakage, didn’t manifest this behavior. Furthermore, the learning curves were not steep, implying that adding more training data did not result in significant improvements in model performance. As for under-sampling, it adversely affected the performance of both classes. This suggests that the information lost during under-sampling is still useful.

Regarding the models we examined, the findings were quite telling. Logistic Regression, a linear model, showcased strong performance, suggesting our data largely follows a linear trend. In contrast, the non-linear models, specifically Random Forest and KNN, tended to overfit and yielded less impressive results. This hints at the absence of intricate patterns in our data that non-linear models might otherwise capture. Thus, for our dataset, a straightforward approach like Logistic Regression seems well-suited.

In essence, images containing only one type of label seem to provide limited value to our model's ability to learn. Instead of turning to more complex models, a promising avenue for improving model performance would be to collect more comprehensively labeled images which featuring both classes.

5.3 Limitations

There are several limitations identified in this research. Firstly, instead of conducting a thorough grid search, hyperparameters were mainly adjusted manually by us. This decision was influenced by the extensive size of the dataset, causing the model training to consume a significant amount of time and resources. It's possible that a more detailed hyperparameter search could uncover better configurations for the model.

Secondly, TRDR data was utilized, not the superior MTRDR data. Most of the noise associated with columns, referred to as instrumental noise, is eliminated in MTRDR. This removal allows bland pixels to be selected not just based on the location of the target pixel, but from anywhere in the image. Techniques like box search might be employed to identify more representative bland pixels. Acquiring more comprehensively labeled MTRDR data might further improve the pixel ratioing approach.

6. Conclusions

Our experimental outcomes affirm the superiority of linear models, such as logistic regression, in handling pixel data for minerals. In contrast to more intricate nonlinear models, linear ones exhibited more stable and efficient performances on our specific dataset, hinting at a pronounced linear relationship amongst the data. This deviates from the prevailing notion that complex models often better capture the nonlinear characteristics of data. Our research offers compelling evidence, showcasing the immense value of simpler models in certain scenarios.

On the matter of data quality and quantity, our investigation further corroborates the pivotal role of comprehensive labeling for model performance. Images containing only a single label offer limited assistance in the learning process for models. This finding accentuates the importance of supplying models with rich, diverse, and thoroughly labeled data, rather than just augmenting the data volume.

The subjects of pixel ratioing and mineral classification continue to be areas of intense interest and challenge. I hope that this research offers a fresh perspective and contributes novel insights to the existing discourse.

Code Resources

All code and experiments are provided and available on GitHub at https://github.com/ese-msc-2022/irp-jw3222.git.

References

Bultel, Benjamin & Quantin, Cathy & Lozac, Loïc. (2015). Description of CoTCAT (Complement to CRISM Analysis Toolkit). IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 1. 10.1109/JSTARS.2015.2405095.

C. Quantin-Nataf, L. Lozac'h, P. Thollot, D. Loizeau, B. Bultel, J. Fernando, P. Allemand, F. Dubuffet, F. Poulet, A. Ody, H. Clenet, C. Leyrat, S. Harrisson, MarsSI: Martian surface data processing information system, Planetary and Space Science, Volume 150, 2018, Pages 157-170, ISSN 0032-0633, https://doi.org/10.1016/j.pss.2017.09.014.

Emanuele Plebani, Bethany L. Ehlmann, Ellen K. Leask, Valerie K. Fox, M. Murat Dundar, A machine learning toolkit for CRISM image analysis, Icarus, Volume 376, 2022, 114849, ISSN 0019-1035, https://doi.org/10.1016/j.icarus.2021.114849.

Murchie, S., et al. (2007), Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) on Mars Reconnaissance Orbiter (MRO), *J. Geophys. Res.*, 112, E05S03, doi:10.1029/2006JE002682.

Viviano, C. E., et al. (2014), Revised CRISM spectral parameters and summary products based on the currently detected mineral diversity on Mars, *J. Geophys. Res. Planets*, 119, 1403–1431, doi:10.1002/2014JE004627.

X. Ceamanos and S. Doute, "Spectral smile correction in CRISM hyperspectral images," *2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, Grenoble, France, 2009, pp. 1-4, doi: 10.1109/WHISPERS.2009.5288992.

Zurek, R. W., and Smrekar, S. E. (2007), An overview of the Mars Reconnaissance Orbiter (MRO) science mission, *J. Geophys. Res.*, 112, E05S01, doi:10.1029/2006JE002701.