Machine-Learning-Driven New Geologic Discoveries at Mars Rover Landing Sites: Jezero Crater and NE Syrtis

Murat Dundar[1], Bethany L Ehlmann[2], and Ellen Leask[2]

[1]Indiana University-Purdue University Indianapolis
[2]California Institute of Technology

November 24, 2022

Abstract

A hierarchical Bayesian classifier is trained at pixel scale with spectral data from the CRISM (Compact Reconnaissance Imaging Spectrometer for Mars) images. Its utility in detecting small exposures of uncommon phases is demonstrated with new geologic discoveries near the Mars-2020 rover landing site. Akaganeite is found in sediments on the Jezero crater floor and in fluvial deposits at NE Syrtis. Jarosite and silica are found on the Jezero crater floor while chlorite-smectite and Al phyllosilicates are found in the Jezero crater walls. These detections point to a multi-stage, multi-chemistry history of water in Jezero crater and the surrounding region and provide new information for guiding the Mars-2020 rover's landed exploration. In particular, the akaganeite, silica, and jarosite in the floor deposits suggest either a later episode of salty, Fe-rich waters that post-date the Jezero crater delta or groundwater alteration of portions of the Jezero crater sedimentary sequence.

Machine-Learning-Driven New Geologic Discoveries at Mars Rover Landing Sites: Jezero Crater and NE Syrtis

Murat Dundar [1], Bethany L. Ehlmann [2,3], Ellen K. Leask [2]

[1]Computer & Information Science Dept., Indiana University-Purdue University, Indianapolis, IN, USA,
[2]Div. of Geological & Planetary Sciences, California Institute of Technology, Pasadena, CA, USA,
[3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Key Points:

- Machine learning can be highly effective in exposing tiny outcrops of uncommon phases in CRISM data
- A new hydrated iron oxide phase, elsewhere on Mars attributed to akageneite, is detected in NE Syrtis and Jezero crater
- Al clays, jarosite, chlorite-smectite, and hydrated silica are reported in Jezero crater

Corresponding author: Murat Dundar, `mdundar@iupui.edu`

Abstract

A hierarchical Bayesian classifier is trained at pixel scale with spectral data from the CRISM (Compact Reconnaissance Imaging Spectrometer for Mars) images. Its utility in detecting small exposures of uncommon phases is demonstrated with new geologic discoveries near the Mars-2020 rover landing site. Akaganeite is found in sediments on the Jezero crater floor and in fluvial deposits at NE Syrtis. Jarosite and silica are found on the Jezero crater floor while chlorite-smectite and Al phyllosilicates are found in the Jezero crater walls. These detections point to a multi-stage, multi-chemistry history of water in Jezero crater and the surrounding region and provide new information for guiding the Mars-2020 rover's landed exploration. In particular, the akaganeite, silica, and jarosite in the floor deposits suggest either a later episode of salty, Fe-rich waters that post-date the Jezero crater delta or groundwater alteration of portions of the Jezero crater sedimentary sequence.

1 Introduction

Hyperspectral data collected by the Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) aboard the Mars Reconnaissance Orbiter have proven instrumental in the discovery of a broad array of aqueous minerals on the surface of Mars since 2006 (Murchie, Mustard, et al., 2009; Pelkey et al., 2007; Viviano-Beck et al., 2014). Although these data have revolutionized our understanding of the planet, existing geologic discoveries are mostly limited to common mineral phases that occur with relatively large spatial extent. Secondary phases on Mars that occur at low abundances are important for a more complete interpretation of the underlying geologic processes. For example, specific minerals such as alunite and jarosite (acidic), serpentine (alkaline, reducing), analcime (alkaline, saline), prehnite (200 °C $<$ temperature $<$ 400 °C), and perhaps phases yet to be discovered, serve as direct environmental indicators of Mars water chemistry. Moreover, the identification of rare phases, even in just a few pixels, enables characterization of the mineral assemblages within a geologic unit, which are critical for identifying the thermodynamic conditions and fluid composition during interactions of rocks with liquid water.

Isolation and discovery of accessory mineral phases is challenging due to the systematic artifacts, random noise, and other limitations of an aging instrument affecting more recently collected CRISM images. The most common spectral mineral-identification method involves ratioing the average spectra from two regions along-track in the image, where the numerator is the spectrum from the area of interest and the denominator is the spectrum derived from a spectrally homogeneous bland region (Carter, Loizeau, Mangold, Poulet, & Bibring, 2015; B. L. Ehlmann et al., 2009; Murchie, Seelos, et al., 2009; Viviano, Moersch, & McSween, 2013). Summary parameters derived from key absorption bands are used to identify candidate regions for the numerator and denominator (Pelkey et al., 2007; Viviano-Beck et al., 2014). Although summary parameters have been effective for detecting common phases with relatively large spatial extent, distinctive absorption bands useful for detecting accessory phases cannot be reliably recovered by summary parameters for two reasons. First, rare phases span a limited number of nearby but not necessarily contiguous pixels in an image, which makes spectral averaging less useful in eliminating random noise. Second, key absorption bands of rare secondary minerals can occur at wavelengths close to those of common phases in the image. The 6.55 *nm* increments between two channels in CRISM offer enough spectral resolution to differentiate between such primary and secondary phases in ideal conditions. However, considering the practical limitations of CRISM data and the occurrence of phases in mixtures, such a distinction may not be possible without exploiting the spectral data in its entirety and identifying less obvious spectral features.

As part of our ongoing efforts to implement machine learning methods to fully automate mineral discovery in CRISM data, we have previously reported new jarosite and alunite detections across Mars (Dundar & Ehlmann, 2016; B. Ehlmann & Dundar, 2015) and have identified a previously unknown CRISM artifact that mimics the characteristics of real mineral absorption at 2.1 μm range that could have significant implications in the search for perchlorate (Leask, Ehlmann, Dundar, Murchie, & Seelos, 2018). Here, we present technical details of our hierarchical Bayesian model and demonstrate its utility by reporting new discoveries of minerals from the NE Syrtis area and Jezero crater and their geologic context. Jezero crater and NE Syrtis are of high interest as regions where the Mars-2020 rover will conduct its in situ exploration and as some of the most dust-free and ancient areas where strata are well-exposed for study of Mars geologic history. Prior studies of Jezero crater and its watershed have focused primarily on the Fe/Mg smectite clays and carbonates that make up deltaic and crater floor deposits and the surrounding, eroded Noachian stratigraphy (B. L. Ehlmann et al., 2008, 2009; Goudge, Mustard, Head, Fassett, & Wiseman, 2015). Here, we focus on identification of small, rare phases to inform the geologic history of the crater in both the crater floor lake sediments, wallrock of Jezero crater, and surrounding region. The region is a well-suited proving ground for the proposed Bayesian model because of its mineral diversity, excellent image availability, and high relevance for Mars exploration.

2 Methods

2.1 Image datasets

We use CRISM I/F data, which are derived by dividing surface radiance by solar irradiance. Radiance data are used for ruling out artifacts during our verification process (Leask et al., 2018). Simple atmospheric and photometric corrections are applied to all images using CRISM Analysis Toolkit (Morgan et al., 2009; Murchie, Seelos, et al., 2009). Only spectral channels that cover the spectral region from 1.0 to 2.6μm (248 channels) are used in this study.

Geographically projected CRISM data were co-registered with high resolution Context camera (CTX) (Malin et al., 2007) and HiRISE (High Resolution Imaging Science Experiment) (McEwen et al., 2007) image data. The CTX global mosaic was used as the basemap for examination of morphology (Dickson, Kerber, Fassett, & Ehlmann, 2018), and standard pipelines for producing local digital elevation models were produced using Caltechs Murray Laboratory pipeline, which utilizes the Ames stereo pipeline (Beyer, Alexandrov, & McMichael, 2018; Shean et al., 2016). CRISM spectral analysis proceeds in multiple steps, described below.

2.2 Creating a training library of spectral patterns by unsupervised learning and visual classification

Over fifty independently characterized CRISM images from the Nili Fossae and Mawrth Vallis regions were processed by a nonparametric Bayesian clustering technique Yerebakan, Rajwa, and Dundar (2014). This method generates a few hundred spectra per image processed, which are visually inspected for mineral detections reported in the literature. Verified spectra are manually classified to create an initial spectral training library. This unsupervised learning approach is not only computationally intensive but also requires a tedious task of manually assigning extracted spectra to classes. Nonetheless, this step is needed to initiate the active machine learning process to collect a representative training library essential for training a robust mineral classifier. In the second phase, the training library collected in this phase is used to implement two models: a bland pixel scoring function for column-wise ratioing and a classifier model that operates on the ratioed data to render mineral classification. Both the scoring function and the classifier uses our two-layer Bayesian Gaussian mixture model.

2.3 Two-layer Bayesian Gaussian Mixture Model

Note that true distributions of spectral patterns in the training library are not known. Different instances of the same pattern detected across different images exhibit varying spectral properties due to differences in atmospheric effects and viewing geometry as well as inherent differences in surface material spectral properties. Our two-layer Gaussian mixture model uses one mixture model for each spectral pattern in the lower layer. Herein, a spectral pattern might represent a mineral phase, a known artifact, a bland pixel category, a common mixed phase, or an unidentified pattern. The number of components in a mixture model for a given pattern is determined by the number of images in which that pattern occurs as the model introduces one Gaussian component for every image the pattern is detected. For example, out of 330 images available in our current training library prehnite exists in eleven of them, which implies that there are elven observed instances of the prehnite pattern ("instance" refers to an occurrence in an image, which can be one or several pixels). The model introduces a Gaussian component for each instance to spectrally model the prehnite pixels corresponding to that instance. Gaussian components corresponding to the same spectral pattern are regulated by a shared local prior and local priors associated with each pattern are in turn modeled by a global prior. In this context the local prior can be thought of as the estimate for the true distribution of the corresponding pattern and the global prior can be interpreted as a template for all viable spectral patterns. This two-layer hierarchical model (illustrated in Figure 1) offers flexibility and robustness for modeling pattern distributions. The lower layer models spectral variations of the same pattern across images whereas the upper layer models spectral variations across patterns. Further technical details of the model and the derivation of the posterior predictive distribution (PPD) is provided in the supplementary material.

2.4 Bland pixel scoring and ratioing

To compute the likelihood of individual pixels originating from bland pattern categories an ensemble version of the model discussed in Section 2.3 is used. Multiple different submodels each with different subset of channels are included in the ensemble. Ensemble models offer better generalizability and are more robust with respect to noise compared to a single model Breiman (2001).

These likelihood scores are then used to identify denominator regions during columnwise ratioing. The denominator is obtained as the average spectrum of a small number of pixels with the highest bland pixel scores sharing the same column as a pixel of interest but lies within a $2w$ row neighborhood of that pixel, where w defines the size of row neighborhood. For robust denominator-insensitive ratioing a range of w values are considered to obtain multiple denominators, and their corresponding ratioed spectra are averaged to obtain a single ratioed spectrum for that pixel. Once all pixels in each I/F image are ratioed this way the ratioed data are used by the pattern classifier for pixel-scale classification.

2.5 Automated pattern classification

Ratioed I/F data are further processed using a cascaded set of one-dimensional median filters with decreasing window sizes to gradually eliminate large spikes Liu, Shah, and Jiang (2004). These ratioed and despiked data are used to train the two-layer Bayesian classifier. This training process involves estimating the parameters of the PPD corresponding to each pattern. Unlike bland pixel scoring, which uses only bland pattern categories, the pattern classifier is implemented with spectral data from all patterns available in the training library. An image is classified at the pixel-scale by evaluating the likelihood of each of its pixel originating from one of the patterns in the training library and then assigning it to the pattern that maximizes this likelihood.

2.6 Active machine learning

The initial training library consisted of patterns detected from a limited number of CRISM images. To obtain a more representative training library, while classifying new images, an active learning scheme is adopted. After each image is classified all detected patterns are visually inspected to confirm automated detections and training library is updated accordingly. The classifier is retrained, i.e., PPDs are updated, every time the training data is updated. The vast majority of images in our training set were selected from Nili Fossae and Syrtis, Mawrth Vallis, Terra Sirenium, Valles Marineris, Libya Montes, and Gale Crater. There are also images processed from elsewhere on Mars to enrich the spectral diversity of detections such as the serpentine detection in Clarites rise, water ice and gypsum detections in polar dunes.

2.7 False positive mitigation

Apart from known artifacts, vertical striping and pixel spiking some CRISM images also suffer from poor signal-to-noise ratio, which is more evident in images acquired at higher temperatures. Some false positives are unavoidable when images are automatically ratioed and classified at pixel-scale. To reduce the number of false positives we use spatial constraints to identify the most viable detections. More specifically, once the image is classified at pixel scale, we map pixel labels onto the image and identify connected components, i.e., groups of pixels sharing the same class label and connected to each other with 8-neighborhood connectivity. All connected components with less than three pixels or all pixels in the same column are considered less viable and are ignored from further processing. We maintain an interactive machine learning workflow to verify all viable detections, especially those with limited spatial exposures. As such, all of the detections reported in this manuscript have been carefully validated by us. Given the numerator region detected by the algorithm, we manually selected a numerator from a similar pixel set and manually selected multiple denominators to verify the pattern identified by the algorithm.

3 Results

3.1 Diverse wallrock minerals at Jezero crater

Mapping of wallrock materials with CRISM data previously revealed low-Ca pyroxenes and Fe/Mg smectites (B. L. Ehlmann et al., 2008, 2009; Goudge et al., 2015). Here we show also Al phyllosilicates and Fe/Mg phyllosilicates, which have an absorption at distinctively longer wavelength than Fe/Mg smectites (Figure 2). The Al phyllosilicates are found on the western crater rim (FRT00005850, HRL000040FF) and the southern crater rim (FRT0001C558) over an elevation range of -2200m to -2500m relative to the Mars datum. The observed Al phyllosilicate spectra have an absorption centered between 2.19-2.20 μm as well as absorptions at 1.4 and 1.9 μm. The slight asymmetry in many of the spectra suggests the presence of kaolinite (Figure 2d). The breadth of some of the Al phyllosilicate absorptions, particularly 1C558 may indicate a mixture of phases, but the breadth is too narrow for opaline silica. The unique Fe/Mg phyllosilicate detections are best isolated right on the rim in FRT0005850 with 1.4, 1.9, and 2.3 μm absorptions. The absorption between 2.32-2.34 μm is longer than that of the Fe/Mg smectites, also observed on the rim (Goudge et al., 2015), and that of the Mg carbonates and Fe/Mg smectites that are common in Jezero crater sediments and basin floor deposits, and this location lacks a 2.5 μm absorption. The spectra are consistent with chlorite or mixed layer Fe/Mg smectite-chlorite phyllosilicates.

3.2 Silica and Jarosite at Jezero crater

As also reported by (Tarnas et al., 2019), we find exposures of hydrated silica within the Jezero crater basin (Figure 2). The exposures have 1.4, 1.9, and 2.2 μm absorptions; the 2.2-μm absorption is substantially wider than in the wallrock Al-phyllosilicates (Figure 2b). At least three small exposures $<500m^2$ are found scattered in the heavily degraded northern delta (FRT000047A3). Locally, the silica is topographically lower and associated with darker, smoother material below the roughened sediments with Fe/Mg smectite and Mg carbonate. These could be confined to a sedimentary bed within the delta, though the orbital data are ambiguous (Figure Suppl. 2i-l) A small exposure of silica is also found on the southernmost lobe of the western delta, adjacent to higher standing carbonate-smectite sediments (HRL000040FF, FRT00005C5E). The exposure is slightly darker in albedo but otherwise unremarkable relative to the surroundings.

In two images (HRL000040FF, FRT00005C5E) material with an absorption of similar width to the hydrated silica is found, but here the band minimum is at 2.26 μm (Figure 2b). This suggests the presence of jarosite, separate or intermixed with the silica, although at the signal to noise of the dataset, mixtures of silica with another mineral cannot be completely excluded. The location and spectral characteristics are the same in both images.

3.3 Akaganeite at Jezero crater and NE Syrtis

A new type of hydrated mineral deposit in Jezero crater was discovered by identifying a cluster of spatially co-located but not always adjacent similar pixels by the hierarchical Bayesian model and then confirmed with traditional ratio techniques (Figure 3). The hydrated phase has a $\sim$1.9-μm absorption that indicates H_2O and a 2.45-μm absorption (Figure 3f). Relative to nearby spectrally "bland" materials there is also a red slope from shorter to longer wavelengths that indicates electronic transitions related to Fe mineralogy different from those of other floor materials. The spectra are most similar to akaganeite $Fe^{3+}(O, OH, Cl)$ and the spectral properties as well as geologic setting near a basin margin are similar to akaganeite reported in Sharp crater (Carter, Viviano-Beck, Loizeau, Bishop, & Le Deit, 2015). The strong 2.45μm absorptions are similar to the 2.42-2.46 μm absorptions found in hydrated and dehydrated akaganeites measured by (Bishop, Murad, & Dyar, 2015; Peretyazhko, Ming, Rampe, Morris, & Agresti, 2018) and are spectrally distinct from the schwertmanite and mixtures of iron oxides/oxyhydroxides measured by these authors. Importantly, the phase is detected in the same locality with the same spectral characteristics in four different images (Figure 3b-3e). The akaganeite-bearing materials are located near eroded remnants of deltas on the Jezero crater floor on the margins of a local topographic low (Figure 3g). The area with akaganeite appears rougher and more rubbly than surrounding floor, with occasional long, cross-cutting ridges (Figure 3g), but is otherwise geomorphologically unremarkable.

Sizeable deposits ($>$0.5 km^2) with an akaganeite spectral signature are also found around NE Syrtis. In CRISM image FRT00019DAA, the signature occurs in basin fill deposits that are incised by a channel that flows west to east over the Syrtis lava flows and is just upstream from late-Hesperian or early Amazonian fill deposits that host Fe/Mg phyllosilicate clay minerals (Figure 4; region further described in (Quinn & Ehlmann, 2019)). The phase is spectrally similar to the akaganeite in Jezero crater but is distinct from nearby polyhydrated sulfate and jarosite spectral signatures (Figure 4d; e.g., (B. L. Ehlmann & Mustard, 2012; Quinn & Ehlmann, 2019). The akaganeite is spatially restricted to a specific deposit on the upstream end of the basin in a local low that erodes into blocky boulders and may exhibit coarse-scale layering on the eastern portion of the outcrop over length scales of 20-50 m (Figure 4c). In addition, north of this location, another deposit of akaganeite in NE Syrtis has been located using the same approach (CRISM FRT00019538), also within small, basin-fill deposits.

4 Discussion

4.1 Two-layer Bayesian Gaussian Mixture Modeling Performance

The proposed hierarchical Bayesian classifier improves mineral mapping in Jezero crater beyond that attained from by-hand work of previous investigators. Small exposures of uncommon phases were identified, testifying to the utility of this approach, which may lead to additional new discoveries elsewhere on Mars and offers new information for interpretation of geologic history.

4.2 Wallrock and Jezero Crater Floor deposits

Jezero crater impacts into a Noachian basement stratigraphy. Fe/Mg phyllosilicates are not unexpected in the wallrock as similar phases are observed in the walls of other impact craters regionally (B. L. Ehlmann et al., 2009; Viviano et al., 2013), specifically Fe/Mg smectites and chlorite. Fe/Mg smectite has been reported previously in Jezero crater (Goudge et al., 2015), and here we show chlorite mixed with smectite is also in the wallrock.

In contrast, Al phyllosilicate has been reported previously on the upper surfaces of the regional Noachian basement, but it is atypical in impact crater walls (B. L. Ehlmann et al., 2009). In Jezero crater, multiple small Al phyllosilicate deposits are associated with the rim region. The detections are ∼2km outside of the crater, right on the rim as well as in down-slumped portions of the rim and discrete blocks (Figure 2; Suppl. Figure 1a-h). The Al phyllosilicates in Jezero crater could result from simple excavation of Noachian basement materials that locally record enhanced alteration. This would be consistent with interpretations of Al phyllosilicate elsewhere in the region. However, except for one coherent block (Suppl. Fig c,d) the occurrences at Jezero crater are associated with materials that surround or embay knobs of excavated rock rather than the rock itself. This could indicate that the Al phyllosilicates along the wall formed from alteration after the Jezero crater impact, in conduits of fluid flow around knobby outcrops, a hypothesis best tested with in situ rover data. Alternatively, the texture of material eroded from the outcrop may enhance the Al-phyllosilicate signal, as interpreted elsewhere on Mars (Wray et al., 2011). Similarly, Al phyllosilicates formed by post-impact alteration or rim rock have been found in situ by the Opportunity rover (Arvidson et al., 2014).

Our finding of silica on Jezero crater floor units expands on similar small exposures reported previously by (Tarnas et al., 2019). The silica may record changes in lake chemistry over time; however, their fairly limited spatial extent, which is not obviously confined to layers, may instead indicate focused zones of groundwater flow and upwelling. Sub-meter scale analysis of rock textures with Mars-2020 will differentiate between these hypotheses.

4.3 Environmental History Implied by Akaganeite

This is the first report of akaganeite in the greater Nili Fossae area. Akaganeite is the best candidate to explain the observed spectral properties of this new phase discovered by the hierarchical Bayesian classifier. Akaganeite forms in Fe-rich, Cl-rich waters, often but not exclusively in acidic environments (Bishop et al., 2015; Peretyazhko et al., 2018); in lab experiments, the acidity promotes crystallinity and sharper 2.46μm absorptions (Peretyazhko et al., 2018).

In both Jezero crater and NE Syrtis, the akaganeite-bearing deposits are associated with eroded, basin-filling materials. The akaganeite setting in local topographic lows is similar to that of the first orbitally-detected akaganeite in Sharp crater (Carter, Viviano-Beck, et al., 2015). The detections in our study area are consistent with a geologic scenario where salty, Cl-bearing, Fe-bearing and possibly acidic Martian waters flowed over the southern Nili Fossae area forming a set of local closed-basin lakes, perhaps dammed by ice (Quinn & Ehlmann, 2019; Skok, Matherne, Karunatillake, & Mustard, 2018). The fluvial activity that formed the NE Syrtis akaganeite is constrained to occur in the late Hesperian or Amazonian by superposition on the Syrtis lavas. At Jezero crater, the setting is more ambiguous as the crater floor unit has been variously attributed to lava or sedimentary fluvial-lacustrine deposits (B. L. Ehlmann et al., 2008; Goudge et al., 2015; Shahrzad et al., 2019). The akaganeite detection is on the margin of a local topographic low in the lake basin where the surface is rubbly and has ridges (Figure 3). In situ rover data are required to determine whether the texture is responsible for the strength of the spectral signature here and whether primary precipitates or groundwater mineralization is responsible. Regardless, the chemistry implied by the akaganeite detections is distinct from the alkaline waters implied by Mg carbonate elsewhere in Jezero crater basin filling sediments. A later episode of salty Fe-/Cl-rich waters during the evaporation of Jezero crater when it was a closed-basin lake is one potential interpretation, to be tested in situ.

4.4 Implications for landed rover exploration

At Jezero crater and NE Syrtis, small detections of uncommon phases are crucial for developing hypotheses about environmental evolution to test in situ, guiding the Mars-2020 rover, and for contextualizing its discoveries. Here we are conservative in our reporting of detections, detailing only those that we were able to verify via traditional techniques, after recognition by the two-layer Bayesian approach. These encompass phases of significance for interpreting the environmental history. However, additional power for operational decision-making about the rover path could come from incorporating all detections and their probabilities into a systematic map of the crater, a potential subject for our future work. Most important is the recognition of possible impact-related alteration (indicated by rim-rock detections) and the changes in Jezero crater lake water chemistry with time implied by the silica and akaganeite.

4.5 The importance of machine learning for planetary hyperspectral data analysis

Our study demonstrates that machine learning can be highly effective in exposing tiny outcrops of specific phases, in CRISM data on Mars that are not uncovered in traditional approaches to imaging spectroscopy data analysis. Here we report results only from Jezero crater and NE Syrtis owing to their significance for upcoming, near-term landed exploration, but similar outcrops of rare phases have been detected across Marsand have the potential to enhance our understanding of Martian geologic history. Moreover, similar techniques can be applied to imaging spectrometer data from other planetary bodies, using machine learning to reveal new insights into planetary processes.

Acknowledgments

Thanks to the CRISM science and operations teams for their work to collect and process these datasets and to Jay Dickson and the Caltech Murray Laboratory for Planetary Visualization for the global CTX mosaic and other assistance with dataset registration. M.D. was sponsored by the National Science Foundation (NSF) under Grant Number IIS-1252648 (CAREER). The content is solely the responsibility of the authors and does not necessarily represent the official views of NSF. E.K.L. was supported by an NSERC PGS-D scholarship. B.L.E. thanks NASA MRO-CRISM extended mission funding for partial support. All CRISM data used in this paper are publicly available through the PDS node (http://ode.rsl.wustl.edu/mars/). Image coordinates of all detections reported in this paper are available as a supplementary file.

References

Arvidson, R., Squyres, S., Bell, J., Catalano, J., Clark, B., Crumpler, L., ... others (2014). Ancient aqueous environments at Endeavour crater, Mars. *Science*, *343*(6169), 1248097.

Beyer, R. A., Alexandrov, O., & McMichael, S. (2018). The Ames stereo pipeline: NASA's open source software for deriving and processing terrain data. *Earth and Space Science*, *5*(9), 537–548.

Bishop, J. L., Murad, E., & Dyar, M. D. (2015). Akaganéite and schwertmannite: Spectral properties and geochemical implications of their possible presence on Mars. *American Mineralogist*, *100*(4), 738–746.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Carter, J., Loizeau, D., Mangold, N., Poulet, F., & Bibring, J.-P. (2015). Widespread surface weathering on early Mars: A case for a warmer and wetter climate. *Icarus*, *248*, 373–382.

Carter, J., Viviano-Beck, C., Loizeau, D., Bishop, J., & Le Deit, L. (2015). Orbital detection and implications of akaganeite on Mars. *Icarus*, *253*, 296–310.

Clark, R., Swayze, G., Wise, R., Livo, E., Hoefen, T., Kokaly, R., & Sutley, S. (2017). *USGS digital spectral library splib07a: U.S. Geological Survey, Digital Data Series 231.* `http://speclab.cr.usgs.gov/spectral.lib06`. (Accessed: 2019-07-03)

Dickson, J., Kerber, L., Fassett, C., & Ehlmann, B. (2018). A global, blended CTX mosaic of Mars with vectorized seam mapping: A new mosaicking pipeline using principles of non-destructive image editing. In *Lunar and Planetary Science Conference* (Vol. 49, pp. 1–2).

Dundar, M., & Ehlmann, B. (2016). Rare jarosite detection in CRISM imagery by non-parametric Bayesian clustering. In *Proceedings of Workshop on Hyperspectral Image and Signal Processing: Evolutions in Remote Sensing (WHISPERS).*

Ehlmann, B., & Dundar, M. (2015). Are Noachian/Hesperian acidic waters key to generating Mars' regional-scale aluminum phyllosilicates? the importance of jarosite co-occurrences with al-phyllosilicate units. In *Lunar and planetary science conference* (Vol. 46, p. 1635).

Ehlmann, B. L., & Mustard, J. F. (2012). An in-situ record of major environmental transitions on early Mars at Northeast Syrtis Major. *Geophysical Research Letters*, *39*(11).

Ehlmann, B. L., Mustard, J. F., Murchie, S. L., Poulet, F., Bishop, J. L., Brown, A. J., ... Wray, J. J. (2008). Orbital identification of carbonate-bearing rocks on Mars. *Science*, *322*(5909), 1828–1832.

Ehlmann, B. L., Mustard, J. F., Swayze, G. A., Clark, R. N., Bishop, J. L., Poulet, F., ... Murchie, S. L. (2009). Identification of hydrated silicate minerals on Mars using MRO-CRISM: Geologic context near Nili Fossae and implications for aqueous alteration. *Journal of Geophysical Research: Planets (1991–2012)*, *114*(E2).

Goudge, T. A., Mustard, J. F., Head, J. W., Fassett, C. I., & Wiseman, S. M. (2015). Assessing the mineralogy of the watershed and fan deposits of the Jezero crater paleolake system, Mars. *Journal of Geophysical Research: Planets*, *120*(4), 775–808.

Leask, E., Ehlmann, B., Dundar, M., Murchie, S., & Seelos, F. (2018). Challenges in the search for perchlorate and other hydrated minerals with 2.1-μm absorptions on Mars. *Geophysical Research Letters*, *45*(22), 12–180.

Liu, H., Shah, S., & Jiang, W. (2004). On-line outlier detection and data cleaning. *Computers & chemical engineering*, *28*(9), 1635–1647.

Malin, M. C., Bell, J. F., Cantor, B. A., Caplinger, M. A., Calvin, W. M., Clancy, R. T., ... others (2007). Context camera investigation on board the Mars reconnaissance orbiter. *Journal of Geophysical Research: Planets*, *112*(E5).

McEwen, A. S., Eliason, E. M., Bergstrom, J. W., Bridges, N. T., Hansen, C. J., Delamere, W. A., ... Weitz, C. M. (2007). Mars reconnaissance orbiter's high resolution imaging science experiment (HiRISE). *Journal of Geophysical Research: Planets*, *112*(E5).

Morgan, F., et al. (2009, March). *CRISM data users' workshop CAT tutorial.* `http://pds-geosciences.wustl.edu/missions/mro/CRISM_Workshop_090322_CAT_MFM.pdf`.

Murchie, S. L., Mustard, J. F., Ehlmann, B. L., Milliken, R. E., Bishop, J. L., McKeown, N. K., ... others (2009). A synthesis of Martian aqueous mineralogy after 1 Mars year of observations from the Mars reconnaissance orbiter. *Journal of Geophysical Research: Planets (1991–2012)*, *114*(E2).

Murchie, S. L., Seelos, F. P., Hash, C. D., Humm, D. C., Malaret, E., McGovern, J. A., ... others (2009). Compact reconnaissance imaging spectrometer for Mars investigation and data set from the Mars reconnaissance orbiter's primary science phase. *Journal of Geophysical Research: Planets (1991–2012)*, *114*(E2).

Pelkey, S., Mustard, J., Murchie, S., Clancy, R., Wolff, M., Smith, M., ... Gondet, B. (2007). CRISM multispectral summary products: Parameterizing mineral diversity on Mars from reflectance. *Journal of Geophysical Research: Planets (1991–2012)*, *112*(E8).

Peretyazhko, T., Ming, D., Rampe, E., Morris, R., & Agresti, D. (2018). Effect of solution ph and chloride concentration on akaganeite precipitation: Implications for akaganeite formation on Mars. *Journal of Geophysical Research: Planets*, *123*(8), 2211–2222.

Quinn, D. P., & Ehlmann, B. L. (2019). A PCA-based framework for determining remotely sensed geological surface orientations and their statistical quality. *Earth and Space Science (Hoboken, Nj)*, *6*(8), 1378.

Shahrzad, S., Kinch, K. M., Goudge, T. A., Fassett, C. I., Needham, D. H., Quantin-Nataf, C., & Knudsen, C. P. (2019). Crater statistics on the dark-toned, mafic floor unit in Jezero Crater, mars. *Geophysical Research Letters*, *46*(5), 2408–2416.

Shean, D. E., Alexandrov, O., Moratto, Z. M., Smith, B. E., Joughin, I. R., Porter, C., & Morin, P. (2016). An automated, open-source pipeline for mass production of digital elevation models (DEMs) from very-high-resolution commercial stereo satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, 101–117.

Skok, J., Matherne, C., Karunatillake, S., & Mustard, J. (2018). The environmental evolution of the NE Syrtis Region. In *Fourth Mars2020 Landing Site Workshop.*

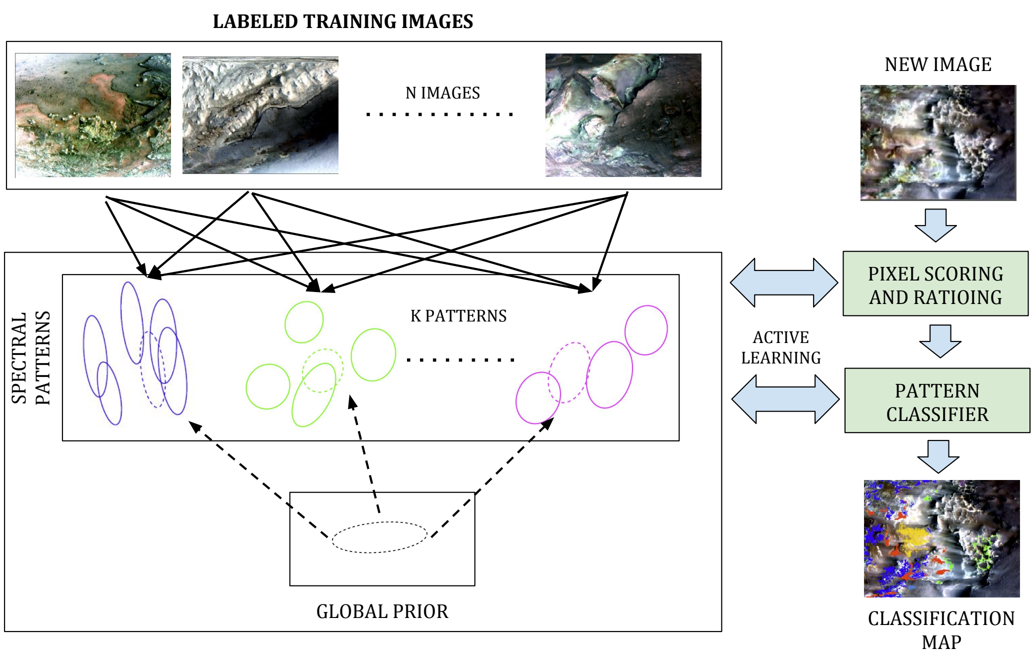
Tarnas, J., Mustard, J., Lin, H., Goudge, T., Amador, E., Bramble, M., & Zhang, X. (2019). Hydrated silica in the Jezero Deltas. In *Lunar and planetary science conference* (Vol. 50).

Viviano, C. E., Moersch, J. E., & McSween, H. Y. (2013). Implications for early hydrothermal environments on Mars through the spectral evidence for carbonation and chloritization reactions in the Nili Fossae region. *Journal of Geophysical Research: Planets*, *118*(9), 1858–1872.

Viviano-Beck, C. E., Seelos, F. P., Murchie, S. L., Kahn, E. G., Seelos, K. D., Taylor, H. W., ... Morgan, M. F. (2014). Revised CRISM spectral parameters and summary products based on the currently detected mineral diversity on Mars. *Journal of Geophysical Research: Planets*, *119*(6), 1403–1431. Retrieved from `http://dx.doi.org/10.1002/2014JE004627` (2014JE004627) doi: 10.1002/2014JE004627

Wray, J., Milliken, R., Dundas, C., Swayze, G., Andrews-Hanna, J., Baldridge, A., ... others (2011). Columbus crater and other possible groundwater-fed paleolakes of Terra Sirenum, Mars. *Journal of Geophysical Research: Planets*,

116(E1).
Yerebakan, H. Z., Rajwa, B., & Dundar, M. (2014). The infinite mixture of infinite Gaussian mixtures. In *Advances in Neural Information Processing Systems (NIPS)* (pp. 28–36).



115 **Figure 1.** Two-layer Bayesian Gaussian Mixture Model Training and Classification

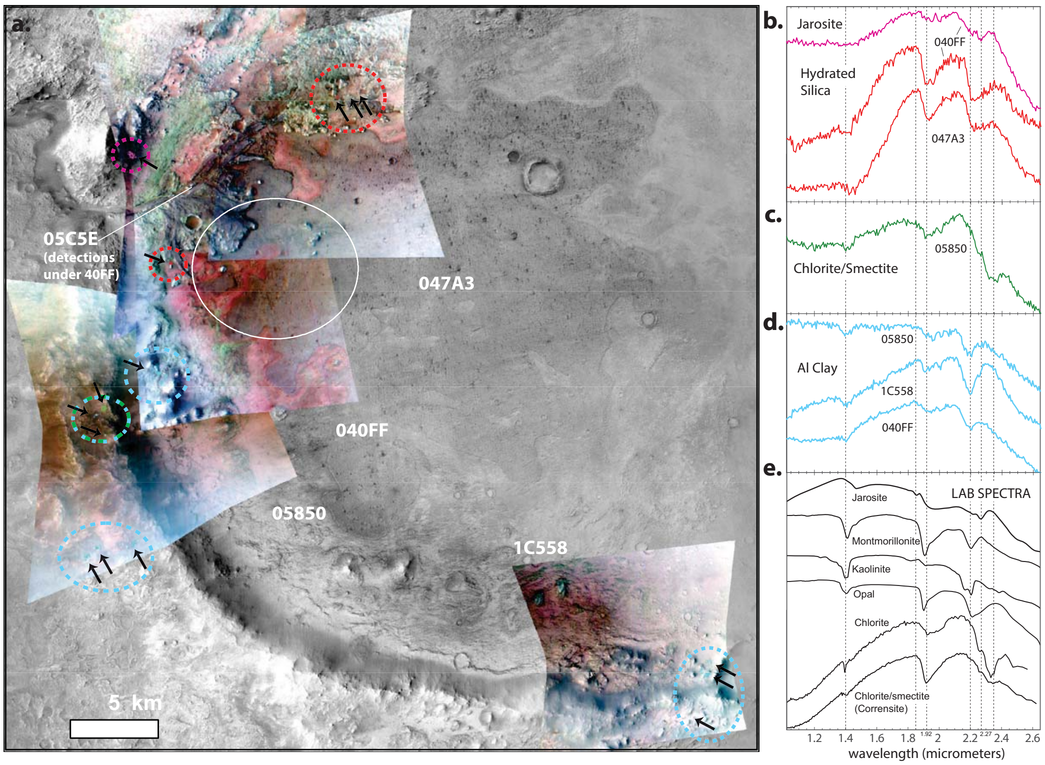


Figure 2. CRISM images of Jezero crater show sub-km exposures of Al phyllosilicates and Fe/Mg phyllosilicates (e.g. corrensite) on the crater walls and hydrated silica and jarosite within basin-filling floor units. (a) CRISM false color images (R: 2.5 μm, G: 1.5 μm, B: 1.1 μm) overlain on a CTX mosaic. The regions of interest with colors corresponding to the spectra in (b-d) are shown, with dashed circles and arrows to flag the locations. Zoom-ins of each area along with corresponding CTX of the same area are shown in the Supplementary Material (Suppl. Figure 1b-d) ratioed CRISM spectra identified by the hierarchical Bayesian algorithm and median filtered to remove spurious spikes (see Supplement for raw numerator and denominator spectra). (e) library spectra from USGS (Clark et al., 2017) and KECK/NASA reflectance experiment laboratory (RELAB).

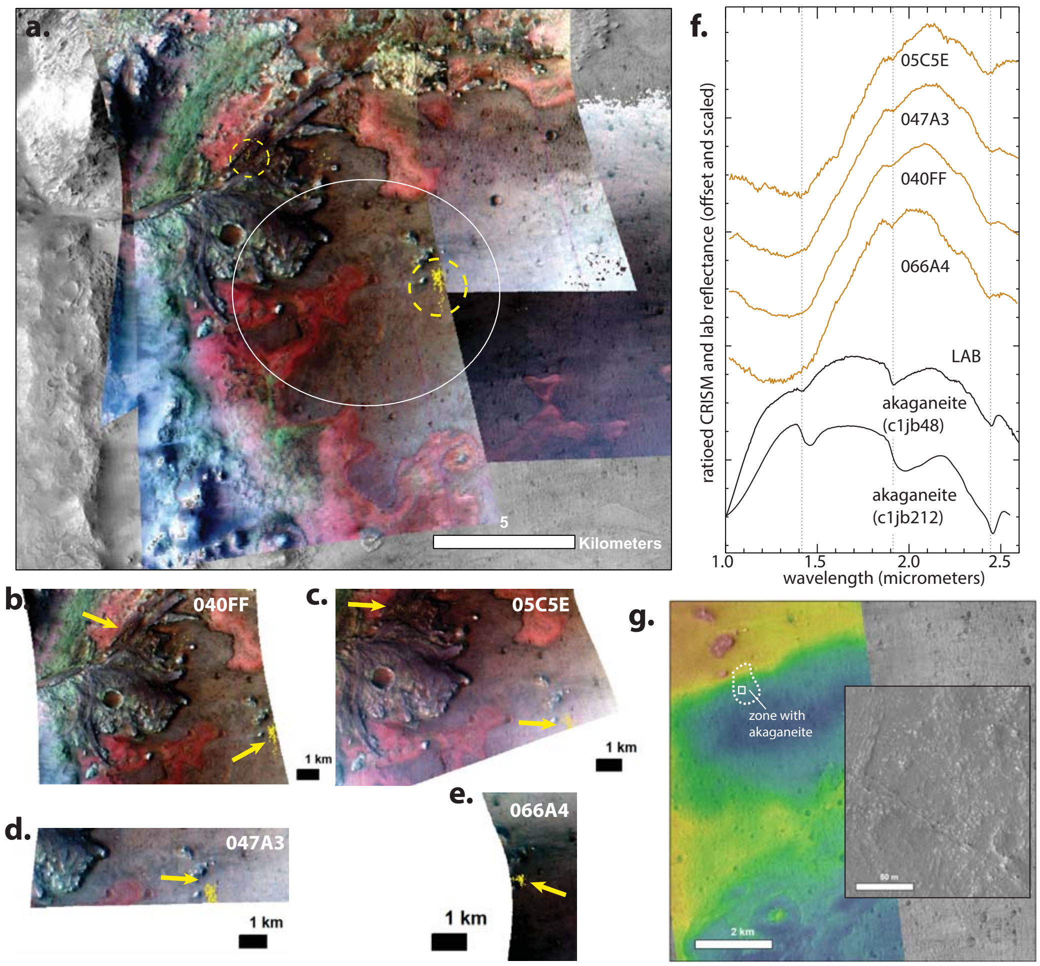


Figure 3. (a) CRISM images covering the floor of Jezero crater show akaganeite. Basemap is the same as Figure 2; yellow regions indicate akaganeite, circled where the pixels are detected in multiple images. (b)-(e) zoom on segments of the CRISM images with the akaganeite sub-km exposures. (f) ratioed CRISM spectra from each of the images compared to laboratory spectra of akaganeite. (g) HiRISE digital elevation model (ESP_023379_1985__ESP_023524_1985) on HiRISE showing the portion of the more rubbly floor materials with akaganeite. Elevations range from -2450 m to -2600 m.

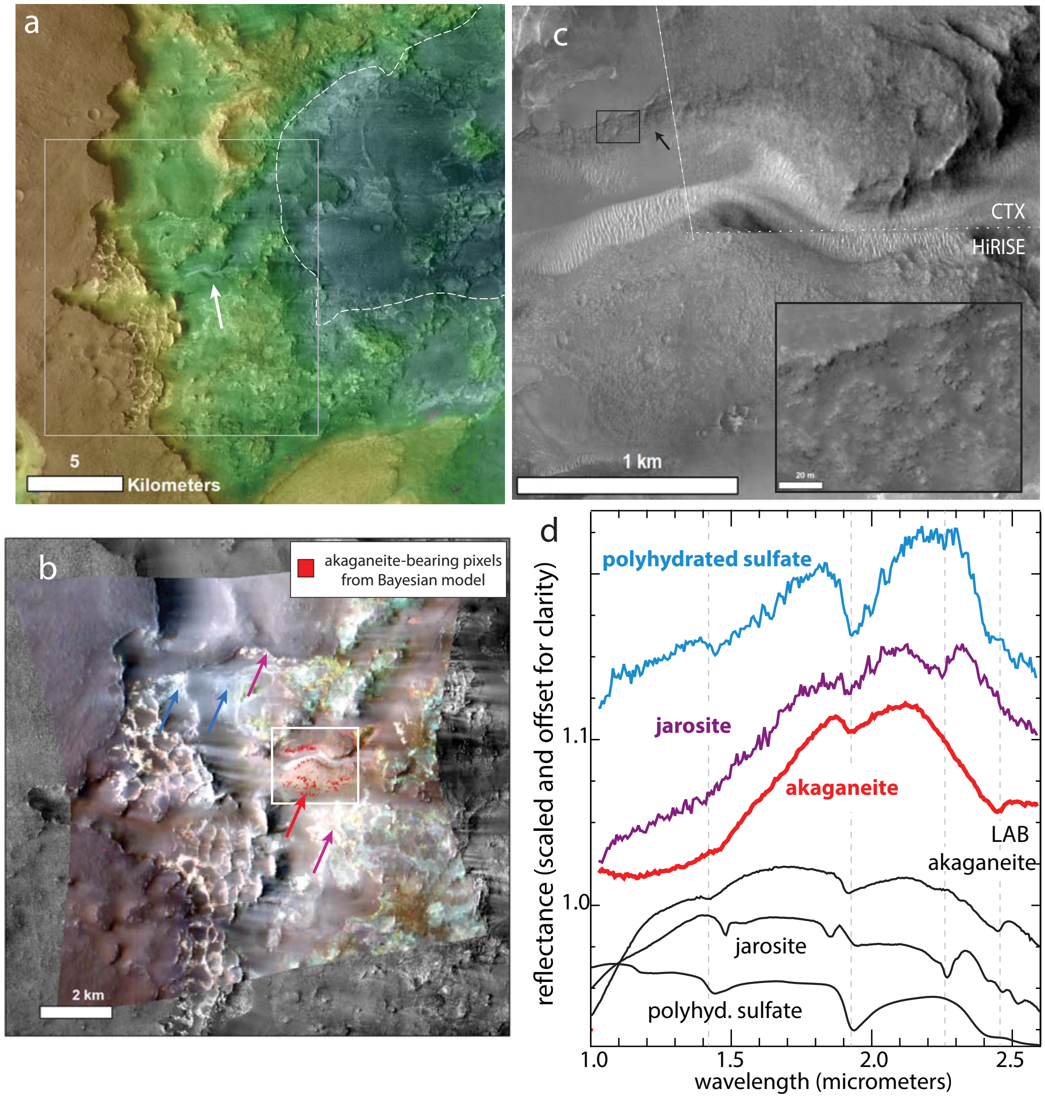
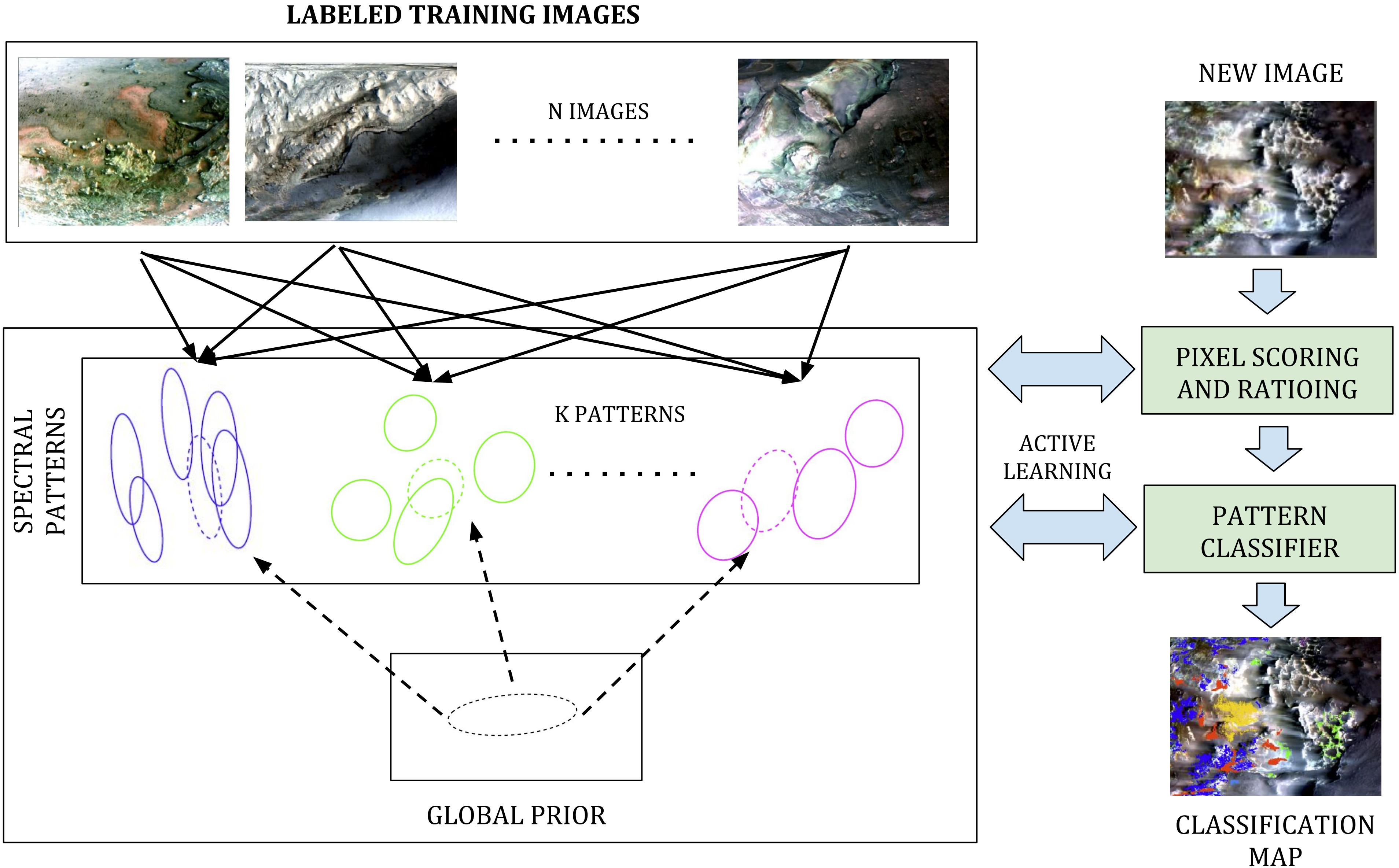


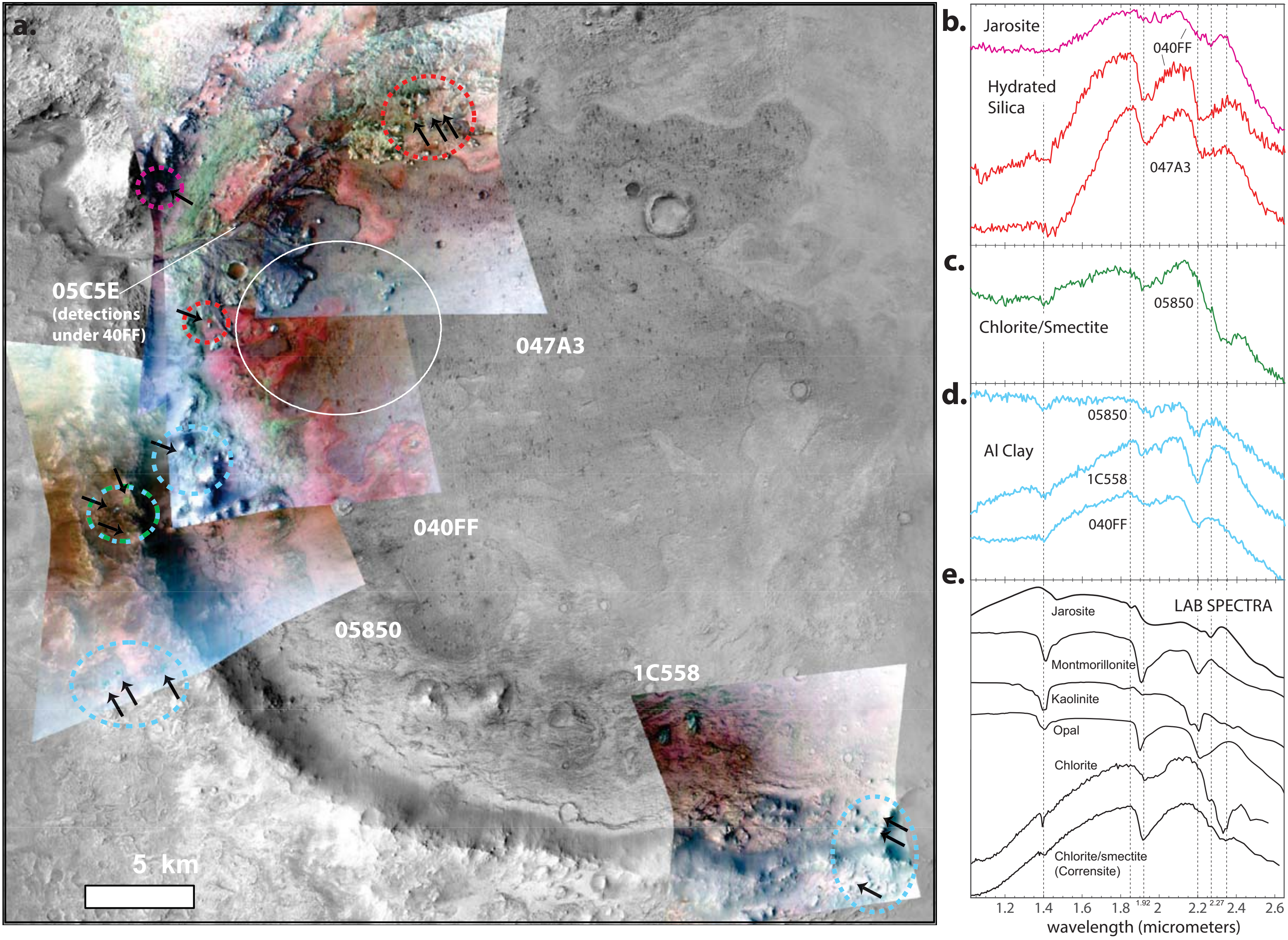
Figure 4. (a) CTX digital elevation model overlapped on a CTX mosaic from (Quinn & Ehlmann, 2019), showing Syrtis lavas and basin-filling deposits, incised by Late Hesperian/Early Amazonian fluvial channels (white arrow). (b) CRISM FRT00019DAA false color image (R: 2.5 μm, G: 1.5 μm, B: 1.1 μm) overlain on the CTX mosaic with pixels of akaganeite detected by a conservative threshold application of the 2-layer Gaussian Bayesian model shown in red. Arrows indicate the approximate locations of the color spectra in panel (d). (c) CTX and HiRISE images of the incised basin-filling deposits, which have the distinctive signature of akaganeite. A black arrow indicates short length-scale, potential coarse-layering or erosion along beds in HIRISE ESP_018065_1975. The inset shows how the deposit erodes into boulders (d) spectra of previously identified polyhydrated sulfates (blue) and jarosite (magenta) from (Quinn & Ehlmann, 2019) along with the new phase we identify as akaganeite (shown in comparison to library spectra in from the RELAB spectral library). The arrows in (B) signify the locations of centers of regions of interest for the spectra.

Figure 1.



LABELED TRAINING IMAGES
N IMAGES
NEW IMAGE
SPECTRAL PATTERNS
K PATTERNS
PIXEL SCORING AND RATIOING
ACTIVE LEARNING
PATTERN CLASSIFIER
GLOBAL PRIOR
CLASSIFICATION MAP

Figure 2.



a.
05C5E
(detections under 40FF)
047A3
040FF
05850
1C558
5 km
b.
Jarosite
040FF
Hydrated Silica
047A3
c.
05850
Chlorite/Smectite
d.
05850
Al Clay
1C558
040FF
e.
Jarosite
LAB SPECTRA
Montmorillonite
Kaolinite
Opal
Chlorite
Chlorite/smectite (Corrensite)
1.2
1.4
1.6
1.8
1.92
2
2.2
2.27
2.4
2.6
wavelength (micrometers)

Figure 3.

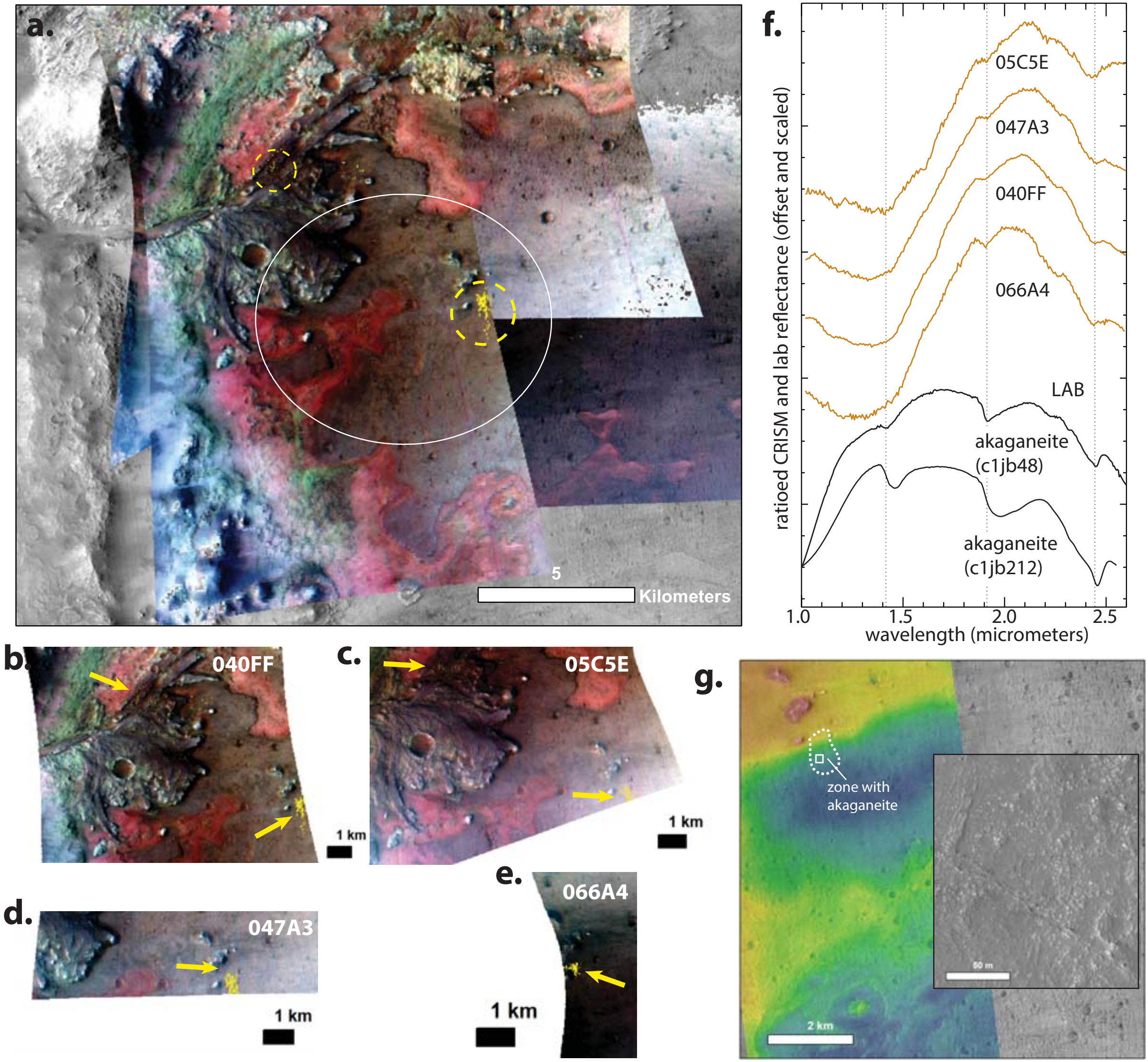
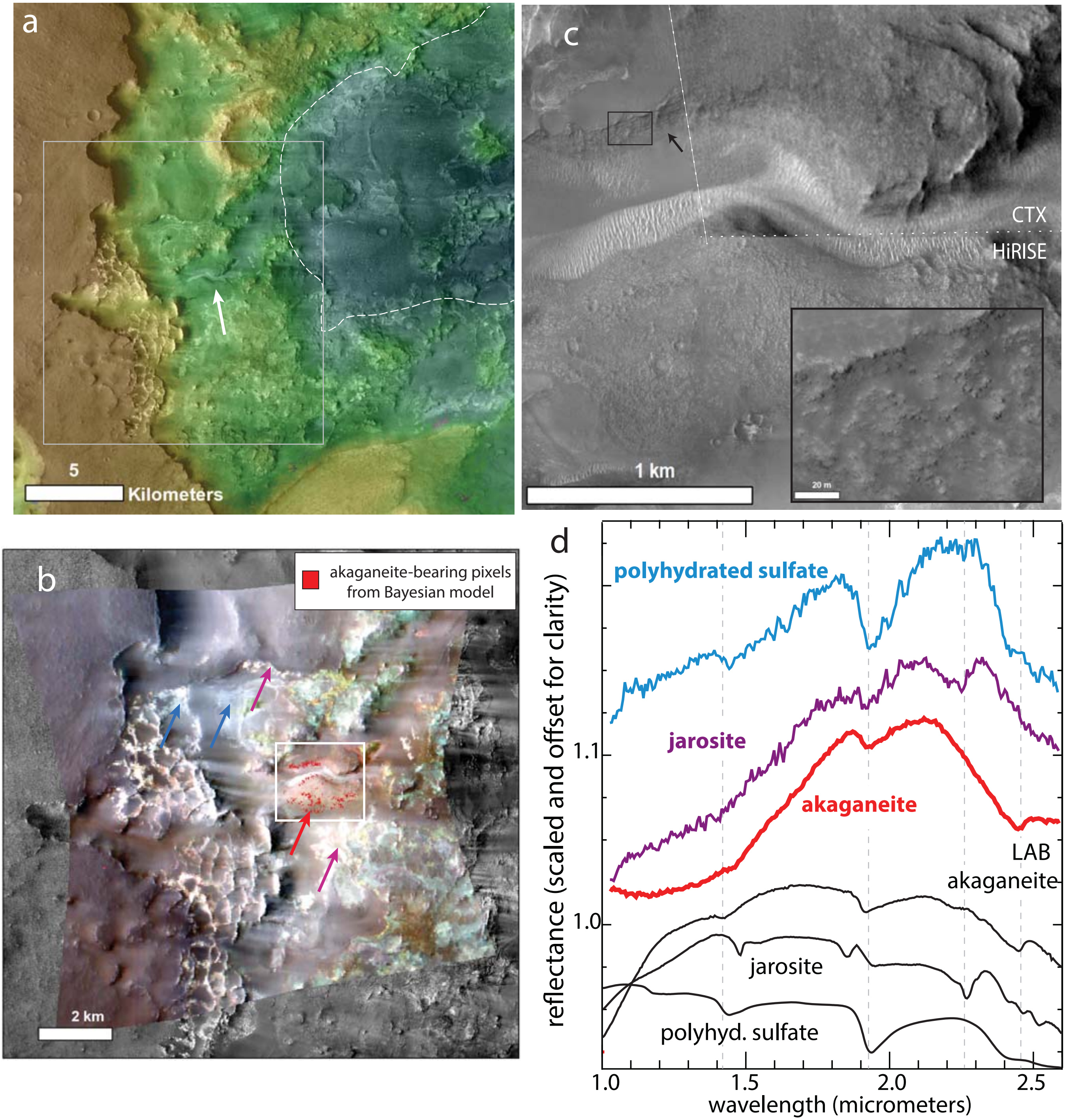

a.
5
Kilometers
f.
05C5E
047A3
040FF
066A4
LAB
akaganeite
(c1jb48)
akaganeite
(c1jb212)
ratioed CRISM and lab reflectance (offset and scaled)
1.0
1.5
2.0
2.5
wavelength (micrometers)
b.
040FF
1 km
c.
05C5E
1 km
d.
047A3
1 km
e.
066A4
1 km
g.
zone with
akaganeite
2 km
50 m

Figure 4.


a
5
Kilometers
c
CTX
HiRISE
1 km
20 m
b
akaganeite-bearing pixels from Bayesian model
2 km
d
polyhydrated sulfate
jarosite
akaganeite
LAB akaganeite
jarosite
polyhyd. sulfate
reflectance (scaled and offset for clarity)
1.1
1.0
1.0
1.5
2.0
2.5
wavelength (micrometers)

Supporting Information for "Machine-Learning-Driven New Geologic Discoveries at Mars Rover Landing Sites: Jezero and NE Syrtis"

Murat Dundar [1], Bethany L. Ehlmann [2,3], Ellen K. Leask [2]

[1]Computer & Information Science Dept., Indiana University-Purdue University, Indianapolis, IN, USA,
[2]Div. of Geological & Planetary Sciences, California Institute of Technology, Pasadena, CA, USA,
[3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Contents

Two-layer Bayesian Gaussian Mixture Model and Derivation of the posterior predictive distribution

We use the following generative model to fit spectral data available in our training library.

$$\text{Data model: } \boldsymbol{x_{ijk}} \sim N(\boldsymbol{\mu_{jk}}, \Sigma_k) \tag{1}$$

$$\text{Local prior: } \boldsymbol{\mu}_{jk} \sim N(\boldsymbol{\mu}_k, \Sigma_k \kappa_1^{-1}) \tag{2}$$

$$\text{Global prior: } \boldsymbol{\mu_k} \sim N(\boldsymbol{\mu_0}, \Sigma_j \kappa_0^{-1}), \Sigma_k \sim W^{-1}(\Sigma_0, m) \tag{3}$$

where k, j, and i are indices used to indicate true patterns, their observed instances, and individual pixels, respectively. $W^{-1}(\Sigma_0, m)$ denotes the inverse Wishart distribution with scale matrix Σ_0 and degrees of freedom m. This model assumes that pixels $\boldsymbol{x_{ijk}}$ are distributed according to a Gaussian distribution with mean $\boldsymbol{\mu_{jk}}$ and covariance matrix Σ_k. Each true pattern is characterized by the parameters $\boldsymbol{\mu_k}$ and Σ_k. The parameter $\boldsymbol{\mu_0}$ is the mean of the Gaussian prior defined over the mean vectors of true patterns, κ_0 is a scaling constant that adjusts the dispersion of the centers of true patterns around $\boldsymbol{\mu_0}$.

Corresponding author: M. Dundar, mdundar@iupui.edu

A smaller value for κ_0 suggests that pattern means are expected to be farther apart from each other whereas a larger value suggests they are expected to be closer. On the other hand, Σ_0 and m dictate the expected shape of the pattern covariance matrices, as under the inverse Wishart distribution assumption the expected covariance is $E(\Sigma|\Sigma_0, m) = \frac{\Sigma_0}{m-d-1}$, where d denotes the number of channels used. The minimum feasible value of m is equal to d+2, and the larger the m is the less individual covariance matrices will deviate from the expected shape. The κ_1 is a scaling constant that adjusts the dispersion of the means of observed pattern instances around the centers of their corresponding true patterns. A larger κ_1 leads to smaller variations in instance means with respect to the means of their corresponding true pattern, suggesting small variations among observed instances of the pattern. On the other hand, a smaller κ_1 dictates larger variations among instances. In Bayesian statistics the likelihood of a pixel $\boldsymbol{x}$ originating from pattern k is obtained by evaluating the posterior predictive distribution (PPD) for pattern k. For our two-layer Gaussian mixture architecture PPDs are derived in the form of *student-t* distributions by integrating out unknown mean vectors and covariance matrices of the true pattern distributions and their observed instances. This directly links observed pattern data with the hyperparameters of the model (κ_0,κ_1,m,μ_0, Σ_0). Optimizing hyperparameters with pixel data from the training library encodes information about observed pattern variations into the model.

Let $\boldsymbol{x}$ be the spectral representation of a pixel in an image to be classified. To classify $\boldsymbol{x}$ we need to evaluate $P(\boldsymbol{x}|\bar{\boldsymbol{x}}_{1k}, \ldots, \bar{\boldsymbol{x}}_{n_k k}, S_{1k}, \ldots, S_{n_k k})$ for each true pattern, where $\bar{\boldsymbol{x}}_{jk}$ and S_{jk} are the sample mean vector and sample covariance matrix of the observed instance j of pattern k. The derivation of $P(\boldsymbol{x}|\bar{\boldsymbol{x}}_{1k}, \ldots, \bar{\boldsymbol{x}}_{n_k k}, S_{1k}, \ldots, S_{n_k k})$ can be carried out in four steps.

In step 1 we integrate out the observed pattern mean vector $\boldsymbol{\mu}_{jk}$ and connect sample mean with the unknown mean vector $\boldsymbol{\mu}_k$ of the true pattern.

$$P(\bar{\boldsymbol{x}}_{jk}|\boldsymbol{\mu}_k, \Sigma_k) \quad = \quad N(\boldsymbol{\mu}_k, \Sigma_k(\frac{1}{n_{jk}} + \frac{1}{\kappa_1})) \tag{4}$$

where n_{jk} is the number of pixels available for observed instance j of true pattern k in the training library.

In step 2 we derive the posterior distribution of the mean vector $\boldsymbol{\mu}_k$ by Bayes rule and show that the posterior mean is weighted average of the sample mean vectors of observed instances and the prior mean.

$$\begin{aligned}
P(\boldsymbol{\mu}_k|\bar{\boldsymbol{x}}_{1k},\ldots,\bar{\boldsymbol{x}}_{n_k k},\boldsymbol{\mu}_0,\Sigma_k,\kappa_0) &= N(\bar{\boldsymbol{\mu}}_k,\bar{\Sigma}_k)\\
\bar{\boldsymbol{\mu}}_k &= \frac{\sum_{j=1}^{n_k}\frac{n_{jk}\kappa_1}{(n_{jk}+\kappa_1)}\bar{\boldsymbol{x}}_{jk}+\kappa_0\boldsymbol{\mu}_0}{\sum_{j=1}^{n_k}\frac{n_{jk}\kappa_1}{(n_{jk}+\kappa_1)}+\kappa_0}\\
\bar{\Sigma}_k &= \bar{\kappa}_k^{-1}\Sigma_k\\
\bar{\kappa}_k &= (\sum_{j=1}^{n_k}\frac{n_{jk}\kappa_1}{(n_{jk}+\kappa_1)}+\kappa_0)
\end{aligned}$$

where n_k is the number of observed instances of pattern k, i.e., the number of training images in which pattern k is detected.

In step 3 we derive the posterior distribution for Σ_k by combining Wishart terms corresponding to all observed instances of pattern k.

$$\begin{aligned}
P(\Sigma_k|S_{1k},\ldots,S_{n_k k}) &= IW(\bar{S}_s,\bar{m}_s) &(5)\\
\bar{S}_s &= \Sigma_0+\sum_{j=1}^{n_k}S_{jk} &(6)\\
\bar{m}_s &= m+\sum_{j=1}^{n_k}(n_{jk}-1)) &(7)
\end{aligned}$$

Finally, in step 4 we derive the posterior predictive distribution for pattern k by integrating out parameters μ_k and Σ_k. Thanks to the conjugacy in our model this operation produces a closed form solution in the form of a *Student-t* distribution.

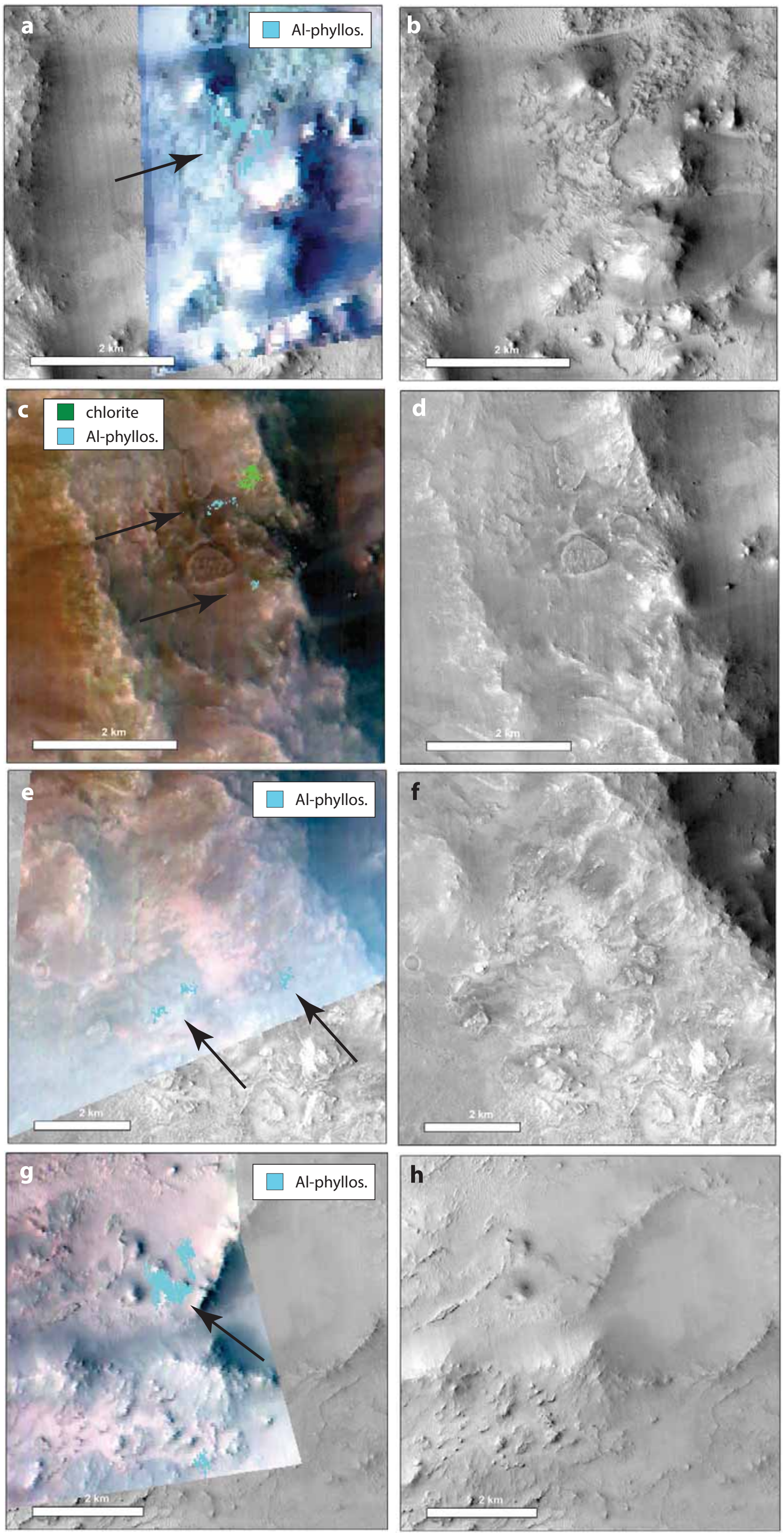
$$\begin{aligned}
P(\boldsymbol{x}|\bar{\boldsymbol{x}}_{1k},\ldots,\bar{\boldsymbol{x}}_{n_k k},S_{1k},\ldots,S_{n_k k}) &= T(\boldsymbol{x}_{ji}|\bar{\boldsymbol{\mu}}_k,\bar{\Sigma}_s,\bar{v}_s) &(8)\\
\bar{\Sigma}_s &= \frac{\bar{S}_s}{\frac{\bar{\kappa}_s v_s}{\bar{\kappa}_s+1}}\\
\bar{\kappa}_s &= \frac{(\sum_{j=1}^{n_k}\frac{n_{jk}\kappa_1}{(n_{jk}+\kappa_1)}+\kappa_0)\kappa_1}{\sum_{j=1}^{n_k}\frac{n_{jk}\kappa_1}{(n_{jk}+\kappa_1)}+\kappa_0+\kappa_1}\\
\bar{v}_s &= m+\sum_{j=1}^{n_k}(n_{jk}-1)-d+1
\end{aligned}$$

Context images of Jezero crater detections

Detections in Jezero crater and the corresponding CTX images are shown in Figures 1 and 2.

Dataset containing image coordinates of automated detections and ratioed spectra

Image coordinates of automated detections reported in this study are provided in the attached file.



64 **Figure 1.** (a) CRISM Al phyllosilicate detections in 040FF with (b) accompanying area in CTX. (c) CRISM Al phyllosilicate and chlorite detections in 05850 with (d) accompanying area in CTX. (e) CRISM Al phyllosilicate detections in 05850 with (f) accompanying area in CTX. (g) CRISM Al phyllosilicate detections in 1C558 with (h) accompanying area in CTX.

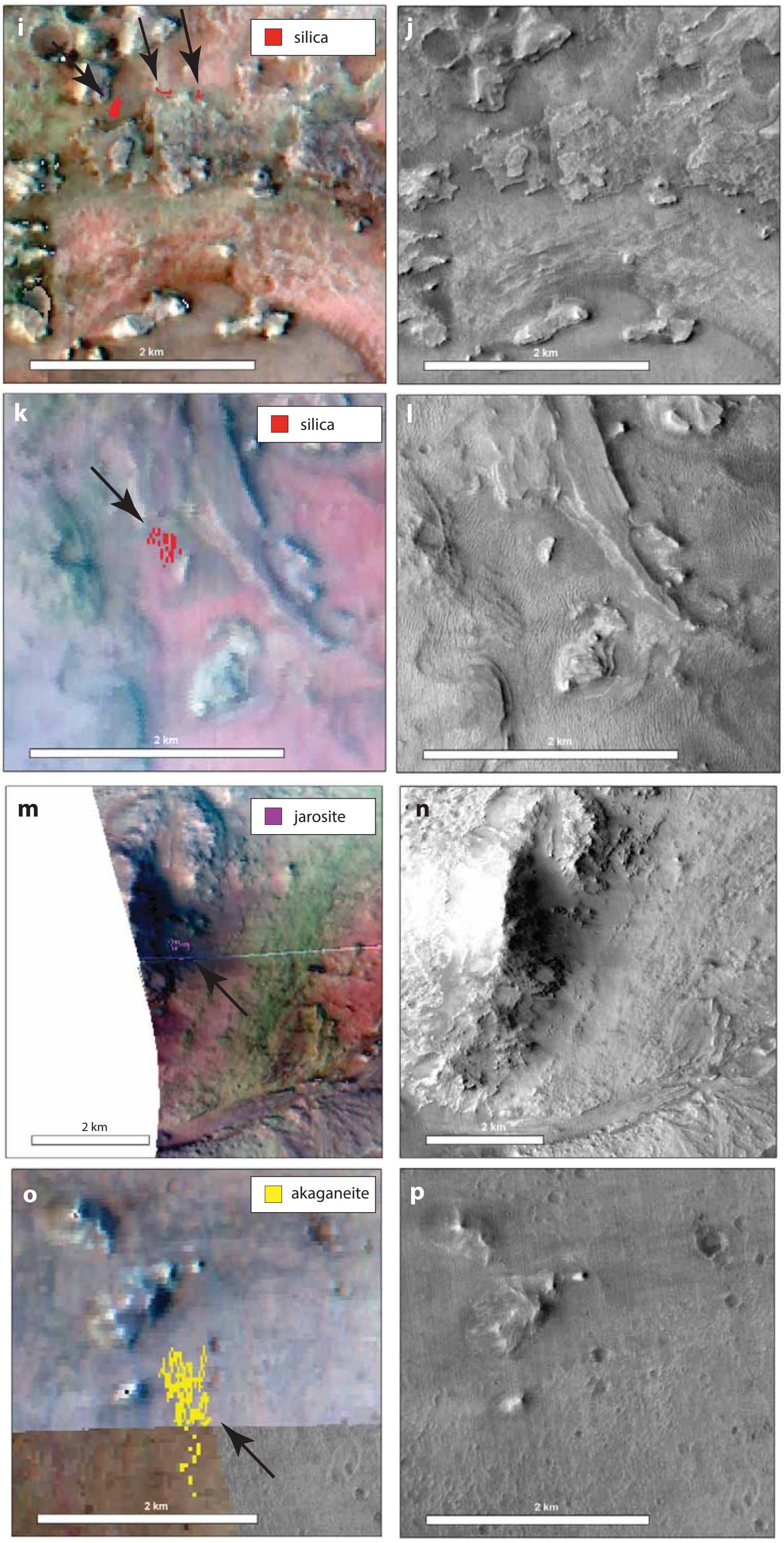


Figure 2. (i) CRISM silica detections in 047A3 with (j) accompanying area in CTX. (k) CRISM silica detection in 05C5E with (l) accompanying area in CTX. (m) CRISM jarosite detection in 05C5E with (n) accompanying area in CTX. (o) CRISM akaganeite detections in 05C5E and 040FF with (p) accompanying area in CTX.