



Science Autonomy and Space Science: Application to the ExoMars Mission

Victoria Da Poian^{1,2*}, Eric Lyness^{1,2}, Ryan Danell³, Xiang Li¹, Bethany Theiling¹, Melissa Trainer¹, Desmond Kaplan⁴ and William Brinckerhoff¹

¹NASA Goddard Space Flight Center, Planetary Environment Laboratory, Greenbelt, MD, United States, ²Microtel LLC, Greenbelt, MD, United States, ³Danell Consulting, Inc., Winterville, NC, United States, ⁴KapScience LLC, Tewksbury, MA, United States

A machine learning approach for analyzing data from the Mars Organic Molecule Analyzer (MOMA) instrument has been developed in order to improve the accuracy and efficiency of this analysis and serves as a case study for the use of machine learning tools for space missions. MOMA is part of the science payload aboard the ExoMars rover, *Rosalind Franklin*, currently planned to land on Mars in 2023. Most NASA robotic space missions return only one thing: data. Remote planetary missions continue to produce more data as mission ambitions and instrument capabilities grow, yet the investigations are still limited by available bandwidth to transmit data back to Earth. To maximize the value of each bit, instruments need to be highly selective about which data are prioritized for return to Earth, as compression and transmission of the full data volume is often not feasible. The fundamental goal is to enable the concept of *science autonomy*, where instruments perform selected onboard science data analyses and then act upon those analyses through self-adjustment and tuning of instrument parameters. In this paper, we discuss the motivations, as well as related work on the use of machine learning for space missions. We also present a first step toward this vision of science autonomy for space science missions. This proof-of-concept exercise for the MOMA instrument aims to develop tools, used on Earth, to support Martian operations of the ExoMars mission. We also discuss the challenges and limitations of this implementation, as well as lessons learned and approaches that could be used for future space science missions.

Keywords: science autonomy, planetary science exploration, decision-making, machine learning, data science, exomars, mars organic molecule analyzer (MOMA), martian operations

INTRODUCTION

A major focus of planetary and small body missions in our solar system is to search for life beyond Earth using a variety of sensors and instruments. Planetary missions have been designed to study the evolution of our solar system and help us understand processes that support habitability and potentially life. These missions face many challenges, including having to withstand extreme environments (e.g., temperatures and harsh radiation) and needing to transmit data great distances back to Earth. For many missions, near Earth or in deep space, high resolution instruments are capable of data acquisition rates that greatly exceed the available bandwidth. Communication bandwidth is currently the main constraint on data-collection practices. As

OPEN ACCESS

Edited by:

Anna Grau Galofre,
Arizona State University, United States

Reviewed by:

Jianguo Yan,
Wuhan University, China
Timothy A. Livengood,
University of Maryland, United States

*Correspondence:

Victoria Da Poian
victoria.dapoian@nasa.gov

Specialty section:

This article was submitted to
Planetary Science,
a section of the journal
Frontiers in Astronomy and Space
Sciences

Received: 04 January 2022

Accepted: 08 April 2022

Published: 28 April 2022

Citation:

Da Poian V, Lyness E, Danell R, Li X,
Theiling B, Trainer M, Kaplan D and
Brinckerhoff W (2022) Science
Autonomy and Space Science:
Application to the ExoMars Mission.
Front. Astron. Space Sci. 9:848669.
doi: 10.3389/fspas.2022.848669

humanity ventures further into the solar system, data becomes increasingly precious while scientific and technical challenges grow. As an example, most current science operations on the Curiosity Rover on Mars, the current state-of-the-art in planetary operations, require manual data review and subsequent decision-making using a ground-in-the-loop workflow, significantly slowing mission activities and tasks. Such operations are not compatible with more distant or extreme (e.g., temperature and radiation) planetary destinations. Science Autonomy—the ability for instruments to autonomously tune, operate, analyze, and direct themselves to optimize science return—is necessary as planetary missions begin to visit these more challenging locations.

Past missions to Venus, Mars, and Titan, as well as several comets and asteroids, have demonstrated the tremendous capabilities of robots for planetary exploration. One example on Mars is the Autonomous Exploration for Gathering Increased Science (AEGIS) system. AEGIS uses machine learning to select sample locations for ChemCam on the Mars Science Laboratory (MSL) based on interpretation of images (Estlin et al., 2014; Francis et al., 2017). Science autonomy has the potential to be as important as robotic autonomy (e.g., autonomously roving terrain) for improving the science potential of these missions because it directly optimizes the returned data. However, on-board science data processing, interpretation, and reaction, as well as prioritization of telemetry, adds new challenges to mission design. As more sophisticated and precise instruments are being developed, the resulting increased data volumes necessitate advanced analysis techniques, such as machine learning (ML), for the full realization of science autonomy.

We first give an overview of some challenges for space exploration missions, then introduce the *science autonomy* concept, and finally describe a use-case when applied to a mission to Mars. We present a first step toward this science autonomy vision: a machine learning approach for analyzing science data from the Mars Organic Molecule Analyzer (MOMA) instrument, which is currently planned to land on Mars within the ExoMars rover, *Rosalind Franklin*, in 2023. MOMA is a dual-ionization source (laser desorption/ionization and gas chromatograph/electron ionization) mass spectrometer that will search for evidence of past and present life on the Martian surface and subsurface through analysis of solid samples (Goesmann et al., 2017; Li et al., 2017). In this case study, we use data collected from the MOMA flight-like engineering test unit (ETU) to develop mass-spectrometry-focused ML techniques that can help scientists perform rapid sample analysis and make optimized decisions regarding subsequent operations. After having described the input data and the overall ML pipeline, we will present first results from the MOMA ML effort, as well as detail the operational deployment and challenges.

Such data characterization and categorization efforts are the first steps toward the longer-term full science autonomy objective → to enable a spacecraft and its instruments to make real-time instrument adjustments during operations, and even prioritize analyses, thus optimizing the complex search for life in our solar system and beyond.

PLANETARY EXPLORATION MISSION CHALLENGES—SCIENCE AUTONOMY CONCEPT

Motivations and Challenges

Extraterrestrial science exploration missions feature key data management challenges, in addition to well-known hardware engineering and operational constraints. Some of these challenges include:

Remote Destinations and Shorter at-Target Mission Lifetimes Limit or Preclude Ground-in-the-Loop Interactions:

Future exploratory missions will travel to more distant targets and/or necessitate shorter lifetimes (e.g., missions on Venus's surface (Gilmore et al., 2020) may last only a few hours, and missions to Europa's surface are expected to be less than 30 days (Hand et al., 2017)). Indeed, the harsh environments these future missions will face, particularly in the realm of extreme temperatures, pressures, or radiation environments, greatly limit how long the mission can be engineered to last. Current successful surface operational models (e.g., Mars, the Moon) are using ground-in-the-loop interactions for supervision and planning to maximize science return, as they benefit from short communication times with Earth, ample power, and/or orbiting satellites as relays.

Remote Destinations and Extreme Environments Present Communication Challenges

Many future missions are planned deeper in the Solar System and/or in extreme environments, involving longer communication delays and smaller data downlink capacities (e.g., up to 25 min one-way light time for missions to Mars, and up to 80 min to reach Saturn). Current mission operations scenarios will not be practical for more distant destinations, whose time-delay is long enough to miss critical science data collection windows.

Remote Destinations Introduce Detection Challenges

As these spacecrafts will operate at greater distances from the Earth, ground-in-the-loop interactions will be limited. Therefore, scientists will be limited in their ability to guide spacecrafts' instrumentation in detecting opportunistic features of interest. When environments or locations of interest can be determined in advance, pre-planned scripts can be used to automate targeting, prioritizing science data analysis during real-time operations. However, for many missions the location of specific features of greatest scientific interest—for instance: meteorite impacts (Daubar et al., 2013), lightning on Saturn (Fischer et al., 2006), icy plumes emitted by Enceladus (Hansen et al., 2006), *etc.*—cannot be predicted, and so the location and timing are unknown.

Data Prioritization Process

Future instruments targeted for destinations beyond Mars (e.g., the Europa Clipper mission) will certainly generate more data

than can be returned to Earth during the nominal mission. The prioritization of data to transmit is vital for these distant exploration missions to optimize mission science return. Ideally, the instruments will run a prioritization process that will, for instance, discriminate calibration data, data below detection limits, and data with instrumental artifacts to identify and preferentially transmit the most compelling data without missing key scientific discoveries.

Related Work

Due to radiation shielding challenges, space missions already suffer from limited computing power (e.g., Random Access Memory (RAM), Central Processing Unit (CPU) limitations). This is a serious concern when considering implementation of advanced ML techniques. However, analytical techniques using ML have previously been deployed in space and demonstrate the possible balance between performance and computational needs. For Earth-based investigations, EO-1 (Earth Observing-One NASA spacecraft) was equipped with software to demonstrate integrated autonomy technologies using the Autonomous Sciencecraft Experiment (ASE) (Chien et al., 2005). This experiment combined a planner (Continuous Activity Scheduling Planning Execution and Replanning (CASPER)) and a Spacecraft Command Language (SCL) system to modify spacecraft activities based on detections (Fischer et al., 2006). The Mars Science Laboratory (MSL) can now autonomously detect targets of interest based on established mission science priorities, thanks to the Autonomous Exploration for Gathering Increased Science (AEGIS) on the ChemCam instrument (Francis et al., 2015). For the landing of Perseverance rover on Mars in February 2021, NASA used artificial intelligence to scan the site prior to landing to select the optimum location for Perseverance to target and safely land on Mars. Moreover, Perseverance is equipped with scientific instruments (i.e., the Planetary Instrument for X-ray Litochemistry (PIXL), Supercam) that use AI and ML during martian operations. Further away, the first investigations at Europa will be realized during flybys of the Europa Clipper spacecraft, planned for launch in the mid-2020's to explore and characterize the habitability of this icy moon (Bayer et al., 2019). Here, onboard detection techniques are under development to enable thermal and compositional anomaly detection, as well as characterization of plumes and icy matter from the subsurface ocean (Wagstaff et al., 2019). Recent studies using multivariate times series for *in situ* detection of interesting events have been tested on the magnetosphere observations collected by the Cassini spacecraft around Saturn (Daigavane et al., 2020), and could be applied to the Plasma Instrument for Magnetic Sounding (PIMS) on the Europa Clipper spacecraft (Westlake et al., 2016). Machine learning and data analysis tools undoubtedly offer great potential to enhance science exploration, but are still in their infancy and can therefore benefit from a comprehensive architecture and operational vision to achieve their full potential.

Long-Term Goal: Science Autonomy

Recently, significant advancements have been made in the development of autonomy for space missions within NASA's

Planetary Science, Heliophysics, Astrophysics and Earth Science divisions (NASA Science, 2018). The recommendations of the 2013–2022 NASA Planetary Decadal Survey (Committee on the Planetary Science Decadal survey, 2011) emphasized the need for mission autonomy development. Most of the recommended developments focused on robotics autonomy. Yet, this last planetary decadal survey did not specifically identify the development of autonomy for science instruments. This autonomy capability would enable quick analysis and information extraction from collected data to focus on the most promising measurements, as its own capability.

Here, we introduce the concept of science autonomy—the ability for science instruments to autonomously tune, operate, analyze, and direct themselves to optimize science return. During real-time operations, the instrument will be able to 1) analyze the collected data to perform self-calibration, 2) prioritize the data downlink based on features detected within the data, 3) adjust operational parameters and 4) make decisions on future operations based on real-time scientific observations.

This science autonomy concept can be split in two categories: instrument *operational automation* and data *interpretational automation*. Operational automation will enhance the data collection process by enabling decisions regarding the instrument's functionality and measurement protocols. This process makes decisions such as selection of sample location, assessment of sample science value, duration of experiments, and selection of instrument analytical parameters; thus, greatly improving the effective sampling rates for *in situ* missions. Instrument automation actions will not only enable science activities in locations where ground-to-space communication scenarios are limited or impossible but will enhance the science return when communication is possible.

The main goal of interpretational automation is to enable data prioritization, and to only send back the most “valuable” data to Earth. Currently, the data collected during missions are processed and interpreted by science and engineering teams on Earth. Not only does this process require significant effort from multiple experts (brain power), but it is a time-consuming task that affects mission operational schedules. The reason for this is that data interpretation is highly specialized. It requires unique expertise as the interpretation is dependent on the instrument, the context of the measurement, and can remain somewhat subjective. While no onboard automation could replace the scientific expertise needed for *full* analysis and interpretation of the returned data, establishing some autonomous pre-processing, including selection algorithms, is an achievable goal that will lead to enhanced science return.

COMPUTER SCIENCE FOR EXOMARS MISSION

Our proof-of-concept data processing technique is intended for use on Earth to support Martian operations during the ExoMars mission. The developed tools are based on data science and machine learning techniques.

ExoMars Mission and Mars Organic Molecule Analyzer Instrument

In the past decades, various missions have been sent to Mars, and while no evidence of life has been found, we know now that this neighbor planet had all the necessary conditions to support life in the distant past (liquid water, a thicker atmosphere, organic material, and other critical nutrients) (Freissinet et al., 2015). However, Mars lost its magnetic field early in its evolution, and thus it has been under constant bombardment by cosmic radiation for at least three billion years. Mars currently has a thin atmosphere (pressure is ~1% of Earth's) that is 96% CO₂, and there is evidence of abundant subsurface ice and possibly liquid water. Studies suggest that if molecular evidence of life is to be found, we should include sampling depth profiles to at least 2 m to reach materials less damaged by radiation (Pavlov et al., 2014).

ExoMars (Exobiology on Mars) is an astrobiology program led by the European Space Agency (ESA) and the Russian Space Agency (Roscosmos). The ExoMars goals (Goesmann et al., 2017) are to search for signs of past and present life on Mars, investigate the water and geochemical environmental variations, and examine Martian atmospheric trace gases and their sources. The 2022 ExoMars Rosalind Franklin rover will carry a drill for martian subsurface sampling to a depth of up to 2 m, and a suite of instruments for geochemistry research.

One of the primary instruments onboard the rover is the Mars Organic Molecule Analyzer (Goesmann et al., 2017): an instrument suite with a primary goal to detect and characterize organic matter within the samples provided by the rover drill and sampling system. The MOMA instrument is centered around a linear ion trap mass spectrometer that is interfaced to two “front-end” techniques: laser desorption/ionization (LDI) and pyrolysis-gas chromatography (py-GC). MOMA represents a notable advance in space mass spectrometry as it is the first to utilize the LDI technique, and the first to feature both tandem mass spectrometry (MS/MS) and selective ion accumulation (SIA) using the stored waveform inverse Fourier transform (SWIFT) technique. MS/MS promises the capability to characterize and potentially identify unknown species detected in a standard mass spectrum *via* analysis of fragment ion masses. Selective ion accumulation is accomplished through the application of a complex multifrequency waveform generated with SWIFT, which allows conversion of a frequency domain profile into a time domain signal. SIA provides improved signal-to-noise ratio (S/N) and limits of detection when analyzing complex samples by focusing the trapping capacity on selected ions of interest.

During martian operations, MOMA will nominally first study a sample using laser desorption mass spectrometry (LDMS) mode, with data returned to Earth for science team analysis and determination/confirmation of the operation to be performed on the next sol (a Mars day) (Goesmann et al., 2017). It will be possible to run further experiments on the same sample, with optionally adjusted parameters (including relocating the laser analysis spot on the sample, or running SWIFT and/or MS/MS), move on to gas chromatography mass spectrometry (GCMS) analysis (with its own options/

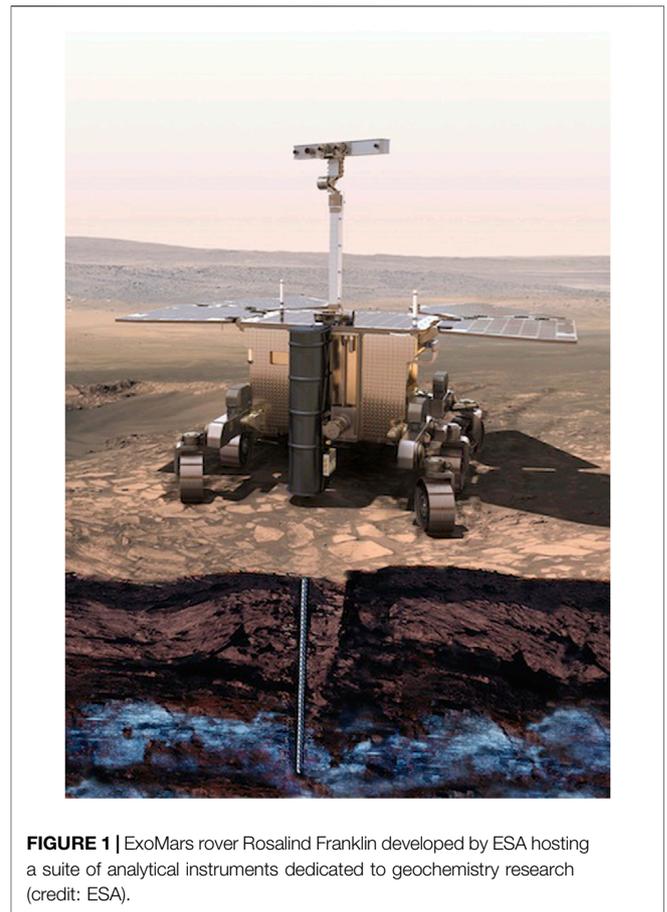


FIGURE 1 | ExoMars rover Rosalind Franklin developed by ESA hosting a suite of analytical instruments dedicated to geochemistry research (credit: ESA).

parameters), or to wait to examine a new sample delivered by the rover's sampling drill. This tactical science planning process is technically challenging. Because of space missions' operations schedules, the team will have as little as 24 h to decide how to further investigate a sample. Anything that could make the analysis more efficient and effective would be beneficial. Scientists need techniques that can quickly extract the information of interest from new data to focus on the salient features for interpretation and tactical planning. Machine learning techniques that utilize the growing body of mass spectrometry knowledge most pertinent to MOMA have the potential to help this time-limited decision-making process during operations, as well as to serve as a proof of concept for inclusion in future instruments and mission designs (Figures 1, 2)

Data Science and Machine Learning Brief Overview

Today, AI is defined as a sub-discipline of computer science that aims at imitating human intelligence. Machine learning is a branch of AI based on the idea that systems can “learn” from data, identify trends and patterns, and make decisions with minimal human intervention. Machine learning can be defined

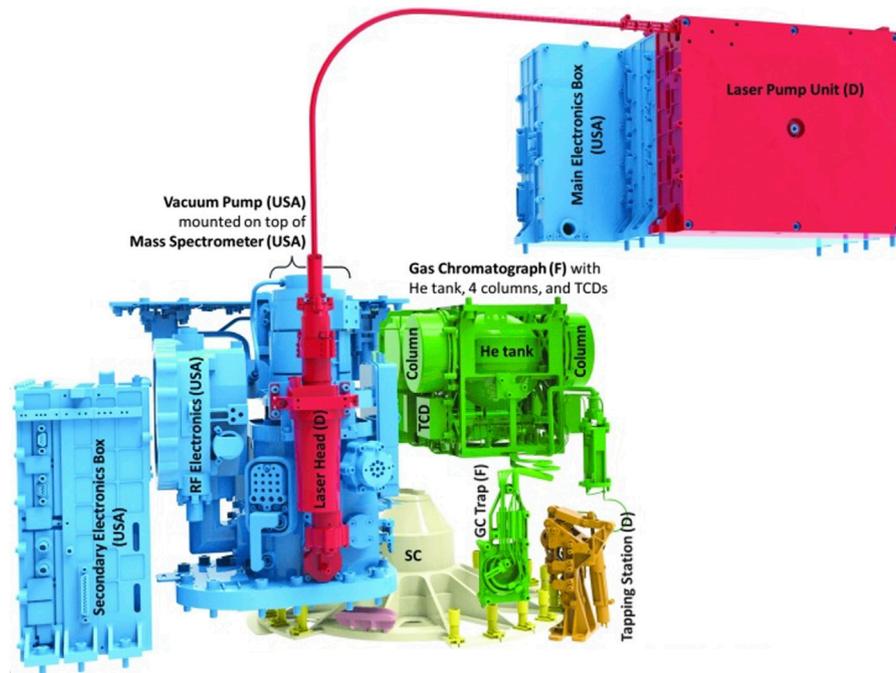


FIGURE 2 | The different parts of MOMA instrument with the country contributing [United States (USA), France (F), Germany (D)] are illustrated. The sample carousel (SC) is composed of 32 ovens (represented in yellow on the figure) for pyrolysis and derivatization. The refillable sample container (holding the sample for LDMS, represented in light pink) and the sample carousel are not part of MOMA (credit: Max-Planck Institute, NASA).

as a method of data analysis that automates analytical model building. The overall goal of ML is to build algorithms (through iterative process) that receive input data, use statistical and mathematical analysis to understand the data and fit that data into models to predict an output. Based on the input data and the task to solve, different types of ML exist. In this project we mainly focus on two types of learning:

- **Supervised:** In supervised learning, algorithms are trained using labeled data (meaning that input and output are known). The algorithms are designed to learn by example to predict the correct output from the input data. The algorithm learns by comparing its results to the correct outputs and modifies the model to minimize the error. The two main objectives are predictive goal (make a prediction of the output label for a new sample) and informative goal (help to understand the relationship between input and output).

- **Unsupervised:** In unsupervised learning, the input data is not labeled, so there are no known correct values of the target variable. The main goal is to explore all the data and study the intrinsic and the structure of the data to either cluster the datasets in similar groups (clustering) or to simplify the datasets (dimensionality reduction). The algorithms learn inherent structure from the input data.

When starting an AI/ML project, a crucial first step is to carefully define the goal such that it is realistic and achievable. Based on the problem to solve and on the available data, the questions, and the strategy to deploy will vary. Moreover, it is essential, during the whole process, to maintain a close collaboration between data scientists and the clients, i.e., end-

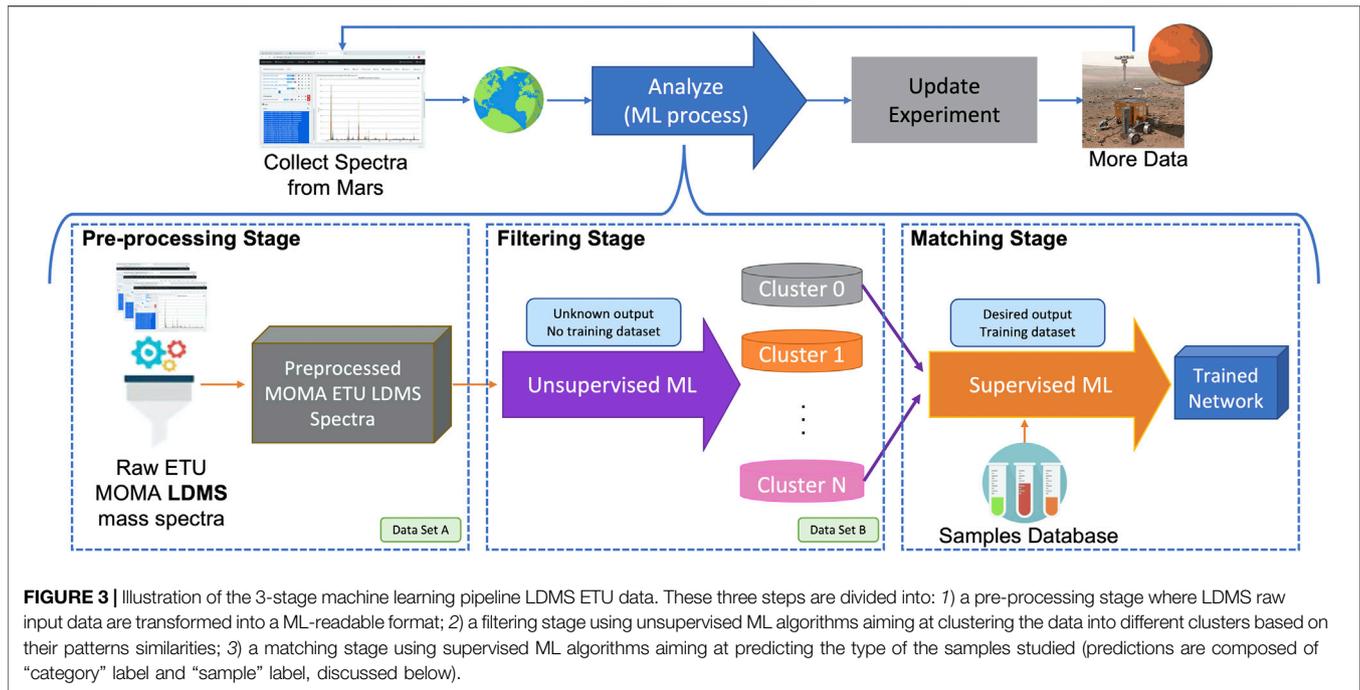
users of the tool or science experts in the field (in our case, mass spectrometry experts). Here, we present the need for ML processing in the analysis of future planetary science data. We mainly focus on mass spectrometry data, but similar techniques could be developed for other planetary science instruments' data.

Machine Learning Development on MOMA Data

The tools developed using MOMA test bed data is a proof-of-concept aiming at supporting the scientists in their decision-making process during martian operations.

Input Data: MOMA ETU and XINA Database

In this study, we used data collected from the MOMA ETU (Engineering Test Unit) in operation at NASA's Goddard Space Flight Center. This flight-analog instrument has been a workhorse that scientists and engineers used to characterize instrument performance and analyze analog samples. During the last several years, over 300 solid samples were tested on the ETU and over 500,000 LDMS spectra were generated. Although MOMA has a GCMS mode and an LDMS mode, for this case study only LDMS data are being studied. In LDMS mode, the laser fires on the surface of the solid sample, desorbing and ionizing atomic and molecular species for mass spectrometry measurements. The analysis of LDMS spectra is notoriously complex and can be time-consuming, thus ML approaches have the potential to supplement existing analysis strategies, making analysis quicker and more accurate. As with



most ML projects, collecting a coherent, well-formatted and complete dataset was the first challenge. The raw data collected from ETU are mass spectra (x-units: m/z , y-units: relative intensity) combined with several experimental parameters.

MOMA maintains a detailed cloud-based database “XINA” that captures every mass spectrum and associated metadata. The metadata have been critical to implementing a working ML model. Firstly, each mass spectrum in the database is associated with a specific sample. However, many of the ETU mass spectra were affiliated with engineering calibration, mechanical checks, or performance tests, and as such are not useful for Mars analog sample classification as their main goal was engineering-oriented rather than science-oriented. Directly admixing these spectra would strongly bias the results and render useful ML impossible. Database queries select the useful science data and eliminate potentially deceptive engineering test data. Our end goal is to develop algorithms will learn from these data to build generalizable models giving classifications -- or find novel patterns -- in new data from Mars of direct use to scientists. This project is structured in three main steps shown in **Figure 3**: data pre-processing, filtering (using unsupervised algorithms), and matching (using supervised training).

Customer-Also-Liked Interface Development

Data Preprocessing

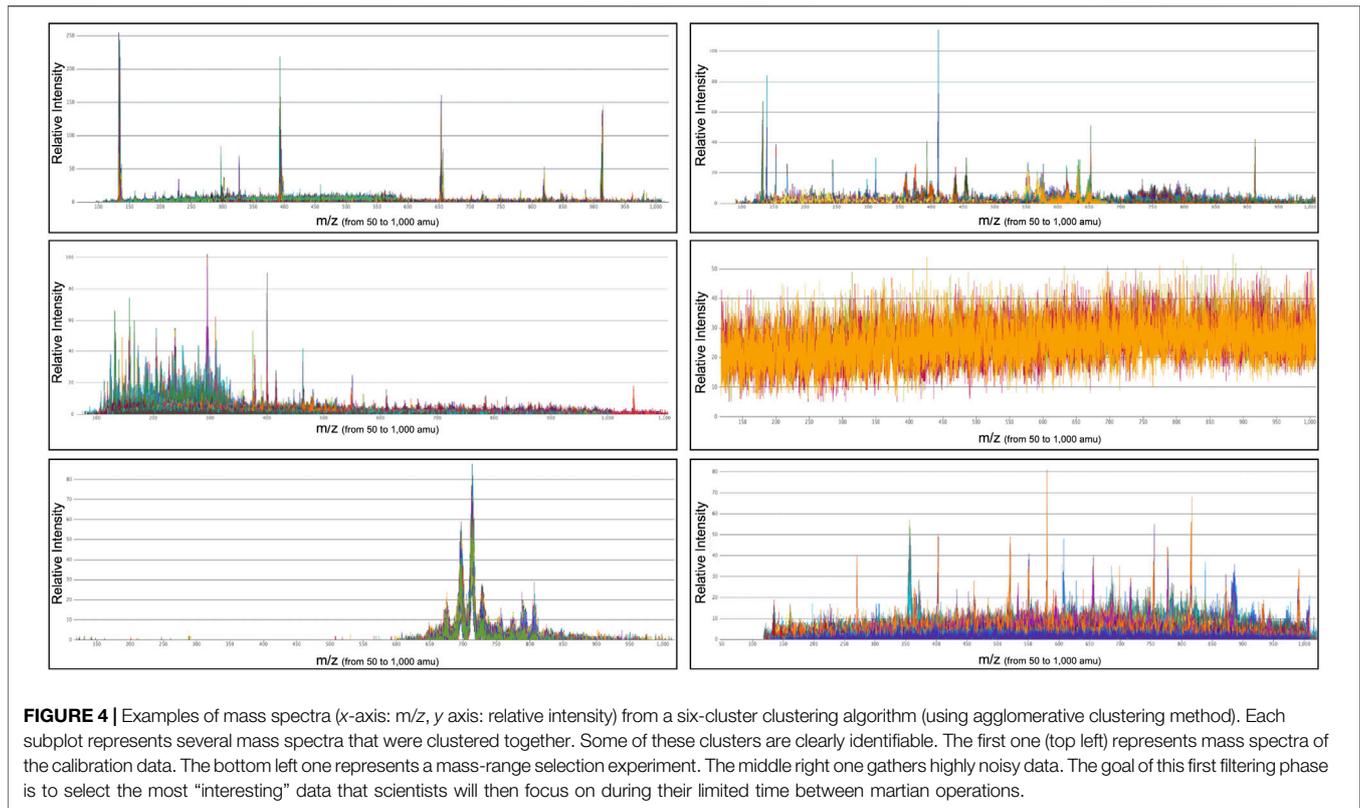
Simple preprocessing “Goldilocks” algorithms, identifying mass spectra with too little or too much signal that degraded spectral quality, enabled removal of many scans. However, there were still many less-pertinent spectra present in the input set from engineering (e.g., mechanical, electrical) tests. We identified and excluded these less-pertinent spectra using querying tools

on spectrum’s metadata. The selected spectra were then transformed into unidimensional arrays by indexing data to rounded, “integerized” (transformed into integers corresponding to the nominal unit mass) values ranging from 1 to 2,000. Selected metadata was then appended to the end of the array for use as ML input data. After completing these data reductions on an original set of more than 500,000 spectra, the pre-processing stage provided a dataset containing 40,000 relevant LDMS mass spectra in a standardized format for further processing.

Filtering Stage

Despite the above-described cultivation, the pre-processed dataset still contained many “misleading” (inapplicable) spectra due to operational errors or test hardware configurations of the ETU. The filtering stage removes these spectra by applying unsupervised learning algorithms from the Scikit-learn (sklearn) library, a ML library available in the Python programming language (Pedregosa et al., 2012). These algorithms have no prior knowledge of the data and simply separate the data into several clusters based on patterns and similarities. Each algorithm has a different pattern-matching method and different input parameters to define the algorithm’s behavior.

An ideal algorithm would separate the applicable “science” data into one set of clusters and the inapplicable data into another set of clusters. To determine which algorithm came closest to this, each algorithm was run against this pre-processed dataset over an array of input parameters, resulting in a two-dimensional search space. To find the optimal separation, two methods were used: mathematical ML performance measures, such as the Davies-Bouldin and Calinski indices, and scientist-based subjective interpretations. These two indices capture both the separation



and the compactness of the clusters. Good clusters should be 1) very compact themselves, and 2) well-spaced from each other. To accomplish the second approach, a series of in-person workshops allowed the MOMA mass spectrometer scientists and the data scientists to sit side-by-side to study the different clustering algorithm outputs, determining which algorithm(s) best clustered the scans according to MOMA science goals. Additional software was written to allow visualization and comparison of the clustering algorithm output.

The most effective algorithms were determined through an iterative process of workshops involving science team feedback and continuously improved algorithm optimization. **Figure 4** illustrates an example output from an algorithm configured to group spectra into 10 clusters. The final output of the filtering stage is a subset of the pre-processed dataset comprising the clusters containing vetted “science” data to be used as a ML training dataset. We will call this dataset the filtered dataset.

This second stage aims at filtering the most interesting data (from a scientific view) from calibration data, noisy data, data with no signal. By clustering similar data together, scientists will be able to focus their attention on unknown and novel data after having checked that the calibration was done correctly and that the instrument was working nominally.

Matching Stage

In the third stage, supervised learning methods were developed using the filtered dataset. Supervised learning starts with training data that are tagged with the “correct answers” or “target values”

where those can be provided. Each mass spectrum is labeled in the database with the name of the sample that the laser targeted to generate the spectra. The first order approach was to take new mass spectra and find the closest mass spectra that the algorithm has been trained to recognize. When a new spectrum is viewed, the software provides a “Customers Also Liked . . .” (CAL) interface that surfaces samples from the database most similar to the spectrum including the quality of the match (this is analogous to what many web-based shopping sites provide to their customers). The algorithm is then further trained on the labelled dataset to identify categories as well as specific samples. After having trained our algorithm, the final model, with a tuned set of weights, would be able to predict answers for similar data that have not already been tagged (i.e., new, and unknown samples analyzed on Mars). New mass spectra fed into our algorithm will return the predicted category, the best match analog sample within the category, and the most similar known scans, providing the scientists all the information about the experiments that generated the known data (instrument parameters and sample information). Several classic supervised algorithms were applied to the tagged dataset. Logistic regression, K-Nearest Neighbors (KNN), decision tree, random forest, Linear Discriminant Analysis (LDA), and MultiLayer Perceptron (MLP) classifier which gave the most accurate results on the input training and testing sets. After this proof-of-concept and based on the superior results from the MLP classifier (a simple neural network (NN) algorithm), we developed customized neural networks models to best fit the MOMA science goals using

TABLE 1 | Performance metrics used for two different NN models on category-labeled and sample-labeled datasets.

	Category		Sample	
	Model 1	Model 2	Model 1	Model 2
Accuracy	0.91057	0.94948	0.89797	0.90929
LogLoss	0.84711	0.91482	0.86699	0.87602
MAE	3.08879	1.74477	3.52376	3.13277
Average PrecisionScore	0.089429	0.05051	0.10202	0.09070

The first model (model 1) is a simple 1-dimensional CNN (convolutional neural network). The second model (model 2) is more complex with more layers implemented. This summary of performance shows that model 2 has a better accuracy in both cases (label being a category or being a sample), nevertheless it is important to do a trade-off between accuracy, other performance, and computation time.

models of neural networks with Keras, an open-source neural network library written in Python. The results of these algorithms are presented below.

EXPERIMENTAL RESULTS AND IMPLEMENTATION

Our streamlined mass spectrometry data processing tool aims at providing two types of decisional information to the scientists. The first, that has been demonstrated and implemented with the methods mentioned in this paper, is to determine if a given spectrum contains data that is similar to other data in the existing dataset. This first step can be used as a preliminary measure of potential scientific relevance and interest. The second type of decisional information is generated after additional training and expert evaluation (currently underway), and will suggest how to tune the instrument to fully analyze the sample's contents. This will help the scientists considerably during the time-limited decision-making process dictated by the Mars surface operations plan. Moreover, this tool could be deployed for other types of spectral data analysis and could expand access to streamlined ML techniques across scientific investigations.

Evaluation of machine learning Techniques

The performance evaluations of our pipeline used two methods: mathematical ML performance measures (example of performance metrics results used for the matching part are listed in **Table 1**) and scientist-based subjective interpretations. The different performance metrics we used to analyze the neural networks results were:

- Accuracy, the count of predictions where the predicted value equals the actual value (the closer to one the better);
- Log loss, the logistic loss measures the performance of a classification model. The loss value mainly represents the prediction error of the neural network, and the goal of the ML model is to minimize this value (the closer to 0 the better);
- Mean absolute error (MAE) is the sum of the “absolute differences” between the predicted and the actual values (the lower the better);

- Average precision score, the average of the precisions (the precision measures how accurate your predictions are, i.e., the percentage of your predictions that are correct) (the closer to 0 the better).

Our preliminary efforts have leveraged collaborative workshops between MOMA mass spectrometer experts and data scientists to analyze different algorithm's performance using custom online visualization tools. The tool we entitled (the) “Customer Also Liked” interface is presented in the following subsection. Overall, these visualization tools enabled the scientists 1) to compare several clustering algorithms and identify the best one for the filtering stage, as well as for the matching stage, 2) visualize most similar spectra to new, unknown spectra and get label predictions to what the analyzed spectra are the most likely to be.

Software Tools Developments

The goal of the matching stage is to identify the most similar known scans to new mass spectra fed into the algorithm. With the prediction of the best match (part of the training dataset), scientists will have access to information about the experiments that generated the known data, such as the instrument settings and the sample information. All this information is summarized in the “Customer Also Liked” interface (**Figure 5**). The user selects a specific dataset (Test ID, TID, rectangle 1) to study as well as specific scans (rectangle 2) and visualizes the generated scan on the top right of the interface (rectangle 3). In the ML section (rectangle 4), the user selects the NN model to use (two models are currently implemented, model 0 and model 1 from the previous table). For each scan, information about the predictions of the category and the sample are shown on the right of the interface (rectangle 5) and the closest match to this selected known scan is shown (rectangle 6). In the example below, scans 211 and 212 of the dataset “8528-BakedHematite” are studied with the NN model 1. The category prediction for the first scan is “Pure Mineral: 94.8%” and the sample prediction is “Hematite: 100%”, both highlighted in green. This information is shown for each scan in the interface.

The different tools on this interface aim to 1) show similar spectra (closest match), and 2) give predictions on the category (i.e., pure mineral in that case) and the sample (i.e., hematite here) classes. Once the scientists get this information, they can go back to the matching data and experiment journal/notes and take that into consideration when deciding what they want to do next with this sample from Mars.

For the longer-term goal of science autonomy, this process of clustering and matching would occur onboard the spacecraft will be followed by a decision regarding the next operations to run. For instance, an autonomous instrument could adjust the experiment settings based on decisions scientists previously made when encountering similar spectra, rather than wait for that decision to come from Earth.

Implementation for Future Missions

Future mission and instrument designs should consider the potential benefit of autonomy techniques using AI (ML, data

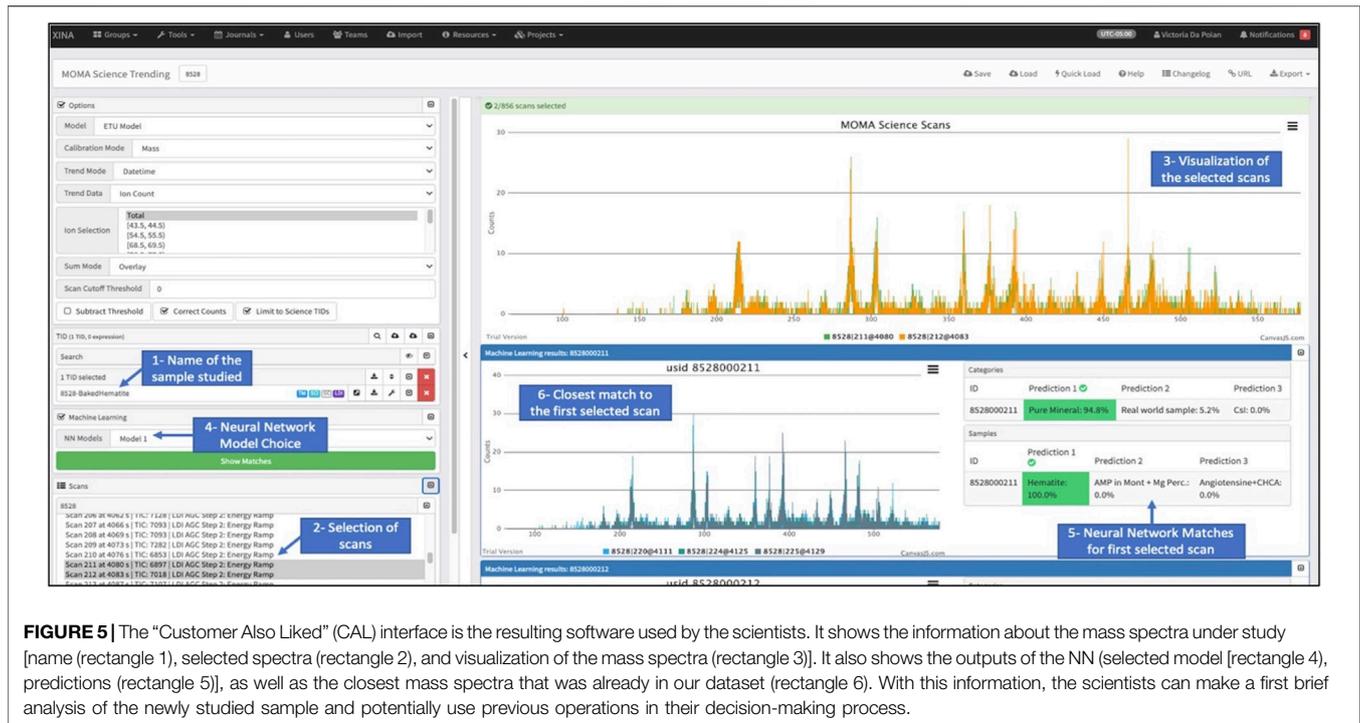


FIGURE 5 | The “Customer Also Liked” (CAL) interface is the resulting software used by the scientists. It shows the information about the mass spectra under study [name (rectangle 1), selected spectra (rectangle 2), and visualization of the mass spectra (rectangle 3)]. It also shows the outputs of the NN (selected model [rectangle 4], predictions (rectangle 5)), as well as the closest mass spectra that was already in our dataset (rectangle 6). With this information, the scientists can make a first brief analysis of the newly studied sample and potentially use previous operations in their decision-making process.

science, etc.). For instance, the Dragonfly mission, a rotorcraft exploring Titan, will be equipped with an instrument similar to MOMA, the Dragonfly Mass Spectrometer (DraMS) (Lorenz et al., 2018) and could greatly benefit from the tools developed during this case study. Mission design stakeholders (engineers, scientists, and mission planners) that choose to utilize AI also need to consider if there are similarities to commercial instruments to take advantage of vast training datasets available, or if additional ground-based instances of flight-like instruments are needed to enhance the data collection process (such as our MOMA ETU). Secondly the mission and instrument architecture need to be designed to support more powerful processors (CPUs) to implement on-board ML. Achieving breakthrough science in extreme environments will likely require strategic instrument control and optimization, necessitating advanced processing.

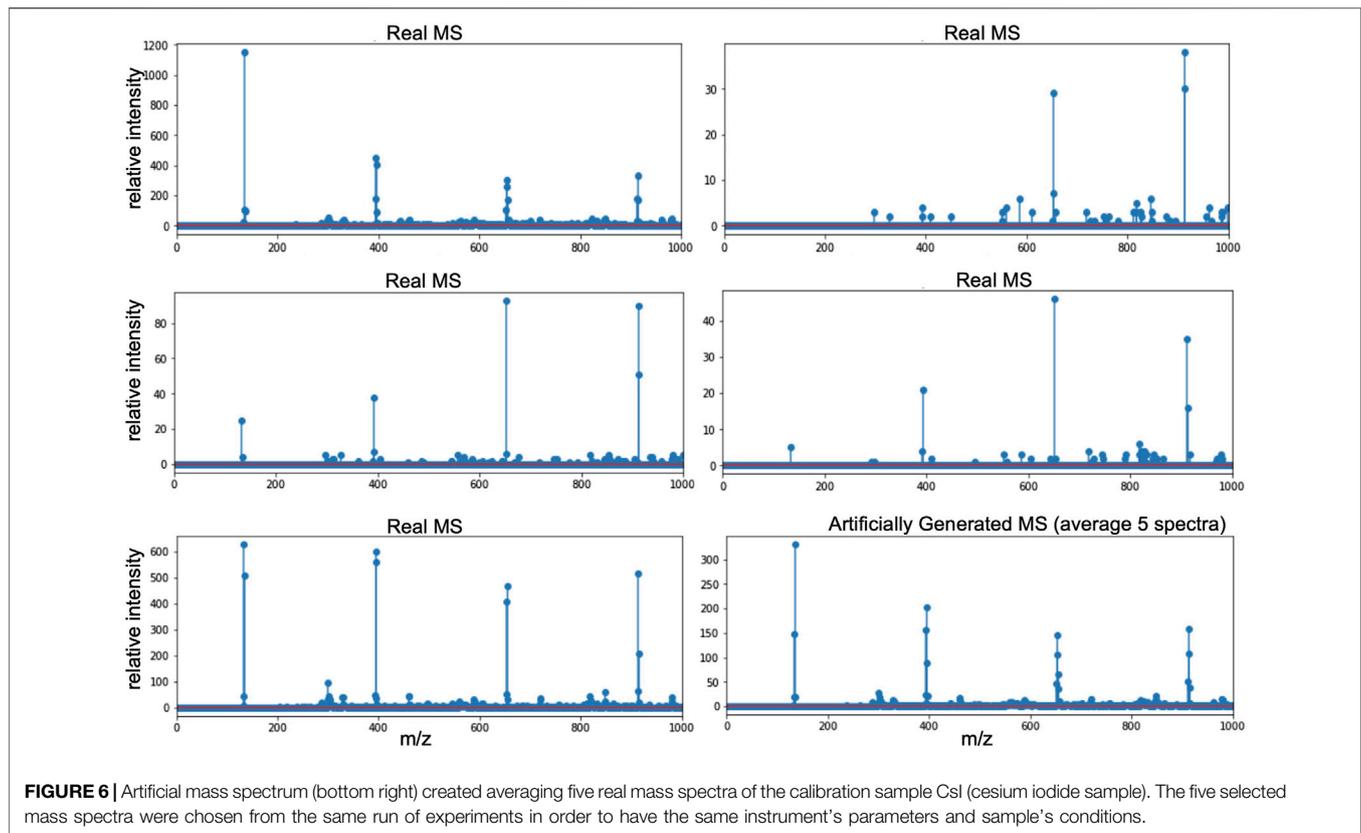
Not only does the mission and instrument design need to accommodate the implementation of ML techniques, but the calibration dataset must also be well defined from the beginning of the mission concept to support the development of effective and accurate algorithms, including “Discovery” detection. Instrument scientists will have to establish a clear sample acquisition strategy based on the science goals of their mission and develop test campaigns with the proper balance of samples in mind. During this project, the data science team had to preprocess data multiple times and reorganize it to remove bias from the large amount of calibration or other unsuitable data. If ML needs are considered and accounted for early in the instrument development cycle, scientists could collect data in a concerted and comprehensive manner, simplifying data science analysis and implementation.

The up-front organization and annotation of the MOMA ETU data made this focused project feasible. However, the actual set of analog samples which have been tested are relatively small and not optimal for direct application of ML approaches. Initial understanding of the range of data expected and the broad categories they would fit in (i.e., data type heterogeneity) is recommended. Research is planned to identify the characteristics of real-world samples that should be analyzed to aid ML algorithms, and to not arbitrarily constrain the planetary applicability of the sample set. Defining clusters and categories in the field of mass spectrometry can be further complicated due to both intrinsic data complexity and subjectivity or bias of expert analysts, which introduces uncertainty into the supervision process. We are investigating ways to remove the subjectivity by employing larger numbers of experts, standardizing expert assignment criteria, and through additional unbiased computational methods.

LIMITATIONS AND CHALLENGES

Data Volumes

As is the case with any AI and ML application, large amounts of data are essential to properly train and tune these powerful techniques. During mission development, scientists use both commercial instruments and flight instrument analog models (such as the Engineering Test Unit, ETU). While all these analog instruments collect spectral data, each instrument’s unique tunable parameters can produce slightly different spectra for identical samples. Moreover, flight spare models, which are designed and built to reproduce the data most accurately from



the flight instrument, are often reserved for mission-critical science activities and thus, the samples that can be tested are very limited. Testing on analog instrumentation can also be constrained due to limited project resources, logistics, sample availability, requirements to stay within hardware operational lifetimes, and complex or time-consuming sample loading and analysis procedures. In the case of MOMA, the amount of data collected with flight analog instruments (i.e., the ETU) is insufficient for more comprehensive ML algorithm development. It is therefore essential to develop methods to appropriately expand our data volume in this and future projects. Data augmentation is a challenging task, as we want to be sure that the generated data will be representative of the analytical mechanisms present and the instrumentation dictating these mechanisms is quite complex.

Artificial Data Generation

The performance of ML models (e.g., neural networks) depend on the quantity and quality of the input data. When data is limited, one approach is to use data augmentation techniques to generate more data as input to the ML models. In short, data augmentation aims to create new, representative, and consistent data by adding slight modifications to the original dataset. Data augmentation is widely used for image classification. Classic image based data augmentation methods include: random rotation, re-scaling, padding, vertical and horizontal flipping, translation (image is moved along x,y direction), cropping, zooming, gray-scaling, adding noise, changing contrast, color modification. Data

augmentation improves model prediction accuracy as it adds more training data into the models, helps resolve class imbalance issues, and reduces the costs of collecting and labeling more data.

In our case, data scarcity is the main reason to look at data augmentation techniques. The different data augmentation algorithms we developed are listed below:

- Average five real mass spectra from the same sample and the same experimental set up (same test ID (TID)) (**Figure 6**).
 - o Input: five real mass spectra from the same sample and same experiment.
 - o Output: 1 mass spectrum averaging these five inputs.
- Add Gaussian noise (**Figure 7**).
 - o Inputs: 1 mass spectrum, noise factor parameter.
 - o Output: 1 mass spectrum with added noise.
- Randomize intensity by adding or subtracting a percentage of the intensity value.
 - o Inputs: 1 mass spectrum, intensity factor parameter.
 - o Output: 1 mass spectrum adding or subtracting a percentage of the intensity based on the intensity parameter (maximum value that can be added/removed).
- Shift all the peaks (of a maximum of 3 AMU) to the left or to the right.
 - o Inputs: 1 mass spectrum, shift factor parameter.
 - o Output: 1 mass spectrum shifting to the left/right all the peaks of the mass spectrum.
- Stretch the mass range (x -axis) of the real mass spectra around the base peak (highest peak).

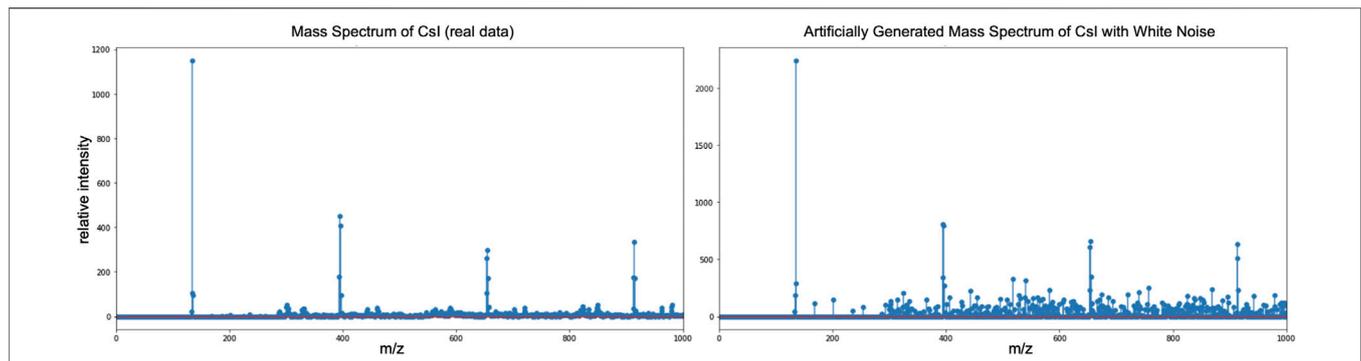


FIGURE 7 | Artificial mass spectrum (right) created by adding Gaussian noise on the real mass spectrum (left). Noise for mass spectrometry experiments can either be from 1) the instrument's noise or from 2) the chemistry noise. For ML training it is often useful to add noise to the input data to reduce overfitting (e.g., when the algorithm is able to learn the training data too well).

- o Inputs: 1 mass spectrum, stretch factor parameter.
- o Output: 1 mass spectrum stretching the mass spectrum around the highest peak (the peaks on the left of the base peak will be shifted to the left and *vice versa*).

By increasing our input datasets using these augmentation techniques, we hope to improve our models' performance. We are currently studying the impacts of combining real and artificial data on our neural networks' algorithms.

Transfer Learning Techniques

Another possible approach to face data volume challenges is to apply "transfer learning" techniques. Transfer learning leverages the knowledge acquired from algorithms developed on large datasets to then tunes those algorithms to adapt them to smaller datasets. In our planetary science study case, we would be looking at leveraging the knowledge gained from commercial instruments used on Earth to develop algorithms for planetary science missions. For example, commercial instruments (such as mass spectrometers) and flight instruments spare models (e.g., ETU models) are instruments that collect chemical spectra data and have many tunable parameters that can produce slightly different results when studying the same sample. Flight spare models are often reserved for mission-related activities and thus commercial instruments are used to collect large amounts of data from relevant samples (e.g., laboratory simulants or Earth-based analogs). Using transfer learning techniques, we will first use these larger commercial instruments datasets to train ML models. Secondly, the learned algorithms will then be tuned and adjusted for the more limited flight instrument dataset ultimately benefiting missions' preparation and operations. This technique is currently being studied and developed on the Sample Analysis at Mars (SAM) datasets. SAM is one of the instruments onboard the rover Curiosity on Mars since 2012. For this study case, we are using commercial GCMS instruments used for years during the pre-mission preparations, the SAM testbed at NASA Goddard Space Flight Center, as well as the planetary data from the SAM

flight instrument on Mars. The results of this proof-of-concept will be published in a following paper.

Technology Limitation

Outer solar system missions present technological challenges along with operational ones. A spacecraft must survive the vacuum of space as well as extreme temperatures changes, preventing the use of common electronics components and requiring advanced thermal management. Also, other harsh environmental conditions, such as the high radiation (e.g., the region around Jupiter) constantly bombarding the spacecraft, can lead to interruption of data collection or downlink, data corruption, and noisy data that will challenge the quality of on-board analysis. All these issues severely constrain the type of hardware and architectures that can be used onboard a spacecraft (e.g., computing power is highly limited). While NASA is working on improving processors for use in high radiation environments, the current state of the art radiation hardened processors are many orders of magnitude slower than processors available for use on Earth. Any space-based algorithm needs to be able to run using lower computing power. Some hybrid designs have been successfully deployed where rad-hard chips provide mission critical functions while faster, but less hardened chips provide data processing functions. Algorithms on these chips would need to be fault tolerant in the case of data upsets or corruption. These environmental factors are being considered during the development of the ML tools and models. However, due to the nuances of each planetary target, additional testing and operational verification will need to be performed once the specific mission architecture is known.

ML Novelty and Trust Issues

Despite all the possible improvements and capabilities ML techniques can offer for space science missions, deploying onboard ML techniques is still considered novel and therefore high-risk. All space agencies are (rightfully so) risk adverse and require that any technology implemented for space use to be fully tested, proven, and vetted. Future missions to distant and more extreme targets will provide fewer opportunities to capture a particular event or measure a sample, and while ML could

provide the flexibility to make such an observation possible, we also need to ensure that it will not lead to false identifications or incorrect conclusions. The reliability of ML techniques will need to be demonstrated *via* well-established testing protocols before being approved for deployment on future missions. Similar to the “Technology Readiness Level” (TRL) used in engineering for hardware maturity classification, an analogous “Trust Readiness Level” system to assess the maturity level of technology using AI/ML tools is proposed and currently being developed. Such a system would utilize a similar structure to the current TRL system where higher levels indicate further verification and validation of the techniques operating under more realistic mission conditions and constraints (i.e., algorithms running on real flight like hardware, software operating autonomously end-to-end and with time limits, and using and processing true unknown data). Properly structuring this testing and making it realistic, particularly in terms of the unknown data to be processed, will be difficult but with vigorous analog studies using available instrument test platforms we believe it will be possible to fully vet and prove these systems are robust and ready for implementation in future missions. The development of this Trust Readiness Level scale is complex as it will not only depend on the target of interest (i.e., Mars, Europa, Venus, *etc.*), but also on the types of instruments [i.e., laser desorption mass spectrometers, gas chromatography mass spectrometers, Laser Induced Breakdown Spectroscopy (LIBS), *etc.*]. This project using LDMS MOMA instrument will be a first step to develop our Trust Readiness Level for LDMS instruments aimed at Mars exploration.

CONCLUSION

In this paper, we introduce some challenges facing future space exploration missions and we emphasize the need for developing science autonomy tools. Upcoming missions, targeting Earth, other planets, the Sun, or deep space, will need greater autonomy. Indeed, by increasing quality science return and reducing the space-to-ground interaction in decision making, more complex and potentially illuminating data collection opportunities will be possible. Our team is investigating data analytics and AI technologies like ML to support onboard data analysis as input for onboard decision-making systems that will generate plans/updates/actions. Increasing the presence and capability of onboard autonomy will redistribute the processes between space and ground to maximize the scientific value of data collected and returned during these constrained missions.

We used MOMA ETU data to develop initial ML algorithms and strategies as a proof of concept and to design software for supporting intelligent operations for autonomous systems. First results of this study show that 1) the preliminary categorization

achieved using the filtering stage could permit autonomous operations such as prioritization of data to be sent to Earth and 2) the prediction made using the matching stage could assist the scientists in their decision-making process during space operations. These initial results illustrate a path where on board processes can make decisions about re-tuning parameters specific for the studied sample, and therefore enable first generation science autonomy. Full science autonomy, which is our ultimate goal, will *minimize* ground-to-space interactions and similarly *maximize* the science return from future missions to challenging locations in our solar system and beyond (Mitchell, 1997).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

VD wrote the first draft of the manuscript. All authors contributed to the writing and editing of the manuscript.

FUNDING

This research project is taking place at NASA Goddard Space Flight Center, in the Planetary Environment Laboratory (code 699) and has been funded by the National Aeronautics and Space Administration (NASA) MOMA science instrument budget. BT was funded through the Internal Science Funding Model (ISFM) Fundamental Laboratory Research (FLaRe) grant at NASA GSFC.

ACKNOWLEDGMENTS

The authors would like to acknowledge the reviewers and the editors for providing feedback which greatly helped in improving the clarity of this manuscript. We thank the MOMA scientists who participated to the different workshops, Andrej Grubisic, Friso Van Amerom, Stephanie Getty, Marco Castillo, Luoth Chou, for their guidance and support for this investigation. We also thank the software team development who supported the project, Joe Cirillo, Nate Lewis, Nick Dobson, Sam Larson, and Joe Avolio. Finally, this project used the performance computing resources from the Science Data Processing division at NASA Goddard Space Flight Center (code 587).

REFERENCES

- Bayer, T., Bittner, M., Buffington, B., Dubos, G., Ferguson, E., Harris, I., et al. (2019). “Europa Clipper Mission: Preliminary Design Report,” in Proceeding of the 2019 IEEE Aerospace Conference, Big Sky, MT, USA, March 2019 (IEEE). doi:10.1109/AERO.2019.8741777
- Chien, S., Sherwood, R., Tran, D., Castano, R., Cichy, B., Davies, A., et al. (2005). “Autonomous Science on the EO-1 Mission,” in Proceeding of the iSAIRAS conference, Nara, Japan, May 2003.

- Committee on the Planetary Science Decadal survey (2011). *Vision and Voyages for Planetary Science in the Decade 2013-2022*. National Academy of Science.
- Daigavane, A., Wagstaff, K., Doran, G. B., Cochran, C., Hackman, C., Rymer, A., et al. (2020). "Detection of Environment Transitions in Time Series Data for Responsive Science," in MileTS '20, San Diego, California, August 24, 2020.
- Daubar, I. J., McEwen, A. S., Byrne, S., Kennedy, M. R., and Ivanov, B. (2013). The Current Martian Cratering Rate. *Icarus* 225 (1), 506–516. doi:10.1016/j.icarus.2013.04.009
- Estlin, T., Gaines, D., Bornstein, B., Schaffer, S., Tompkins, V., Thompson, D. R., et al. (2014). "Automated Targeting for the MSL Rover ChemCam Spectrometer," in 12th International Symposium on Artificial Intelligence, Robotics, and Automation in Space (i-SAIRAS).
- Fischer, G., Desch, M., Zarka, P., Kaiser, M., Gurnett, D., Kurth, W., et al. (2006). Saturn Lightning Recorded by Cassini/RPWS in 2004. *Icarus* 183 (1), 135–152. doi:10.1016/j.icarus.2006.02.010
- Francis, R., Estlin, T., Doran, G., Johnstone, S., Gaines, D., Verma, V., et al. (2017). AEGIS Autonomous Targeting for ChemCam on Mars Science Laboratory: Deployment and Results of Initial Science Team Use. *Sci. Robot.* 2. doi:10.1126/scirobotics.aan4582
- Francis, R., Estlin, T., Gaines, D., Bornstein, B., Schaffer, S., Verma, V., et al. (2015). "AEGIS Autonomous Targeting for the Curiosity Rover's ChemCam Instrument," in Proceeding of the 2015 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, Oct. 2015 (IEEE). doi:10.1109/AIPR.2015.7444544
- Freissinet, C., Glavin, D. P., Mahaffy, P. R., Miller, K. E., Eigenbrode, J. L., Summons, R. E., et al. (2015). Organic Molecules in the Sheepbed Mudstone, Gale Crater, Mars. *J. Geophys. Res. Planets* 120, 495–514. doi:10.1002/2014JE004737
- Gilmore, M. S., Beauchamp, P. M., Lynch, R., and Amato, M. J. (2020). "Venus Flagship Mission Decadal Study Final Report," in *A Planetary Mission Concept Study Report Presented to the Planetary and Astrobiology Decadal Survey*.
- Goesmann, F., Brinckerhoff, W. B., Raulin, F., Goetz, W., Danell, R. M., Getty, S. A., et al. (2017). The Mars Organic Molecule Analyzer (MOMA) Instrument: Characterization of Organic Material in Martian Sediments. *Astrobiology* 17, 655–685. doi:10.1089/ast.2016.1551
- Hand, K. P., Murray, A. E., Garvin, J. B., Brinckerhoff, W. B., Christner, B. C., Edgett, K. S., et al. (2017). *Europa Lander Study 2016 Report – Europa Lander Mission*.
- Hansen, C. J., Esposito, L., Stewart, A. I. F., Colwell, J., Hendrix, A., Pryor, W., et al. (2006). Enceladus' Water Vapor Plume. *Science* 311, 1422–1425. doi:10.1126/science.1121254
- Li, X., Danell, R. M., Pinnick, V. T., Grubisic, A., van Amerom, F., Arevalo, R. D., et al. (2017). Mars Organic Molecule Analyzer (MOMA) Laser Desorption/ionization Source Design and Performance Characterization. *Int. J. Mass Spectrom.* 422, 177–187. doi:10.1016/j.ijms.2017.03.010
- Lorenz, R. D., Turtle, E. P., Barnes, J. W., Trainer, M. G., Adams, D. S., Hibbard, K. E., et al. (2018). Dragonfly: a Rotorcraft Lander Concept for Scientific Exploration at Titan. *Johns Hopkins APL Tech. Dig.* 34, 3.
- Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.
- Nasa Science (2018). *2018 Workshop on Autonomy for Future NASA Science Missions: Output and Results*.
- Pavlov, A. A., Eigenbrode, J., Glavin, D. P., and Floyd, M. (2014). "Rapid Degradation of the Organic Molecules in Martian Surface Rocks Due to Exposure to Cosmic Rays Severe Implications to the Search of the "Extinct" Life on Mars," in Proceedings of the 45th Lunar and Planetary Science Conference, Feb 2014.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., and Thirion, B. (2012). Scikit-Learn: Machine Learning in Python. *J. Machine Learn. Res.* 12, 2825–2830.
- Wagstaff, K., Doran, G., Davies, A., Anwar, S., Chakraborty, S., Cameron, M., et al. (2019). "Enabling Onboard Detection of Events of Scientific Interest for the Europa Clipper Spacecraft," in KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, July 2019, 2191–2201. doi:10.1145/3292500.3330656
- Westlake, J. H., Mcnutt, R. L., Kasper, J. C., Case, A. W., Grey, M. P., Kim, C. K., et al. (2016). *The Plasma Instrument for Magnetic Sounding (PIMS) on the Europa Clipper Mission*. American Astronomical Society. DPS meeting #48, id 123.27.

Conflict of Interest: VD and EL were employed by Microtel LLC. RD was employed by Danell Consulting, Inc. DK was employed by KapScience LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Da Poian, Lyness, Danell, Li, Theiling, Trainer, Kaplan and Brinckerhoff. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

GLOSSARY

AEGIS autonomous exploration for gathering increased science

AI artificial intelligence

ASE autonomous spacecraft experiment

CAL customer also liked

CASPER continuous activity scheduling planning execution and replanning

CPU central processing unit

DraMS Dragonfly Mass Spectrometer

EO-1 Earth observing-one

ESA european space agency

ETU engineering test unit

ExoMars exobiology on Mars

GCMS gas chromatography mass spectrometry

KNN K-Nearest Neighbors

LDA Linear Discriminant Analysis

LDI laser desorption/ionization

LDMS laser desorption mass spectrometry

LIBS laser induced breakdown spectroscopy

MAE mean absolute error

ML machine learning

MLP MultiLayer Perceptron

MOMA Mars organic molecule analyzer

MSL Mars science laboratory

MS/MS tandem mass spectrometry

NN neural network

PIMS plasma instrument for magnetic sounding

PIXL planetary instrument for X-ray litochemistry

py-GC pyrolysis-gas chromatography

RAM random access memory

SAM sample analysis at Mars

SCL spacecraft command language

SIA selective ion accumulation

sklearn Scikit-Learn

S/N signal-to-noise ratio

SWIFT stored waveform inverse Fourier transform

TID test ID

TRL Technology Readiness Level