Optimized Martian Dust Displacement Detection Using Explainable Machine Learning

Ana Lomashvili* Kristin Rammelkamp* Olivier Gasnault† Protim Bhattacharjee*

Elise Clavé* Christoph H. Egerland* Susanne Schröder* Begüm Demir‡ Nina L. Lanza§

Abstract

The ChemCam instrument on the Curiosity rover performs geochemical analyses of rocks on Mars using Laser-Induced Breakdown Spectroscopy (LIBS). The shockwaves generated during the LIBS measurements sometimes shift dust from the surface of the target. The study of the Martian dust phenomena in the scope of the ChemCam instrument has the potential to provide insight into the planet's geology and aid calibration methods for data processing. In this study, we develop a pipeline, named Dust Displacement Detection (DDD), for automatic detection of dust displacement on LIBS targets based on the image dataset acquired by ChemCam. To this end, we introduce a data preprocessing methodology and test two-stage models with a pretrained model in the first stage for feature extraction and a Random Forest classifier or a Support Vector Machine as a binary classifier in the second stage. The best performing model was found to consist of the first 10 layers of VGG16 and a Random Forest classifier, achieving 92% accuracy. Additionally, we use Explainable AI (XAI) methods such as Shapley values and guided backpropagation for model optimization. The experiments show potential for model optimization, and the application examples presented encourage discussion of machine learning in the field of Martian dust research.

1. Introduction

In 2012, NASA's Curiosity rover landed on Mars to investigate the Martian surface, geology, and climate, with a particular focus on assessing the planet's past habitability. The rover has successfully completed its primary mission, finding signs of past habitability. After many years of service, Curiosity is still at work and is now on its fourth extended mission [16, 34, 37]. The rover is equipped with various scientific instruments including the mast-mounted ChemCam (Chemistry and Camera) instrument. ChemCam consists of the first Laser-induced Breakdown Spectrometer (LIBS) in planetary science and a Remote Micro Imager (RMI) [12–14]. LIBS involves focusing a laser on the surface of the target up to several meters from the rover, creating a luminous micro-plasma that emits characteristic photons from excited atoms, ions, and molecules. Spectral analysis of this light provides spectra with emission lines from species present in the sample, from which elemental composition can be derived [6]. The RMI instrument images through the same telescope as the LIBS providing context to the samples. RMI images are usually taken before and after the LIBS measurements [14]. Until now, the instrument has acquired data from more than 4000 individual targets and collected LIBS spectra from multiple points of each target (5-25 points per target) typically arranged in rasters. The LIBS plasma is accompanied by a shock wave that expands into the thin Martian atmosphere which sometimes leads to dust displacement on the target [22]. The occurrence of dust displacement is often observed in the images taken after the LIBS measurements, as shown in Fig. 1. The left and right images correspond to the RMIs acquired before and after LIBS measurements, respectively; in the "after" image, dust displacement is marked by an ellipse and the LIBS pits are circled. Investigating whether dust displacement has occurred and whether it is related to specific rock types or local or seasonal conditions on Mars, can support the investigation of the Martian surface and its atmosphere. The Martian dust itself is also of interest to scientists, as it provides insights into the planet's atmospheric processes and rock cycles [11, 35]. Dust displacement properties can be used to characterize dust in terms of thickness and, together with LIBS observations, in terms of geochemical composition. In addition, the study of dust displacement can provide valuable information for the understanding of the laser-

*German Aerospace Center (DLR)
†Institut de Recherche en Astrophysique et Planétologie (IRAP)
‡Technical University Berlin
§Los Alamos National Laboratory

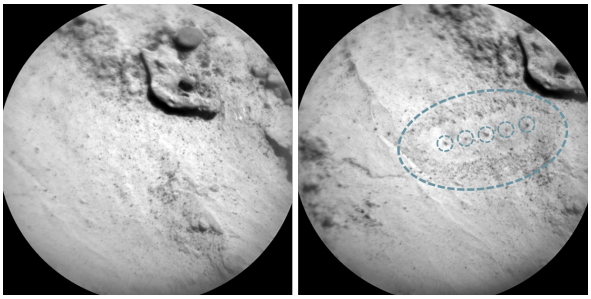

Figure 1. RMI images taken before (left) and after (right) LIBS measurements. The post-LIBS image displays small craters caused by multiple LIBS ablations and a contour of dust displacement around them, dashed lines highlight those contours. Target *Lacropte* measured on sol 3146.

induced plasma on Mars and its accompanying phenomena. For example, the shock wave responsible for the dust displacement could serve as an indicator of the quality of the laser-target coupling [4, 29]. Such supplementary information can contribute to the development of new LIBS data normalization approaches.

This work focuses on optimally detecting whether dust displacement took place during ChemCam LIBS measurements based solely on RMI images. In this study, a limited number of samples are manually labeled, preprocessed, and trained on a machine learning pipeline employing pretrained convolutional neural network and Random Forest classifier. Additionally, explainable AI methods are used to explore model optimization. The optimal automated detection allows efficient derivation of seasonal or site-specific dependencies of dust displacement along the traverse, potentially providing insight into Martian processes. The study motivates further investigation of properties, such as the shape of the dust displacement or its application to other Mars rovers.

2. Literature review

2.1. Martian dust and ChemCam RMI investigations

Martian dust has been previously investigated due to its strong influence on the atmosphere's temperature and sedimentary processes [26]. One example is the research presented in [11]. They examined ChemCam LIBS data of the first laser shot only and used multivariate regression to retrieve its elemental composition [5]. The study explored whether dust contributed to an amorphous component of soil. The work was based on LIBS data and the use of RMI images showing dust displacement was not considered. In another work, ChemCam RMI images were explored and classified texture-wise using deep learning techniques [9]. The main challenge that they encountered was a limited number of labeled data, which often occurs in real-world machine learning problems. This problem was addressed by using the pretrained model Inception V3 [10]. Such an approach is also known as *transfer learning* [20] and the procedure is as follows: a model is trained on a huge dataset with millions of images such as ImageNet [7]. The weights of the CNN filters that best describe the images are learned during the training process. Later, the trained network can be repurposed for another classification problem with different data and categories. Using this method, the authors of the study classified LIBS targets into nine different classes with an accuracy of 80%. Another ChemCam target classification based only on image data was performed by automatic partitioning of weakly connected components [8]. The work showed the variability of LIBS targets and an effective way to group them.

2.2. Explainable AI

Over the years, developments in the field of machine learning have led to more complex models. With the complexity of models comes the difficulty of understanding what they are learning and how they are making predictions, turning them into so-called "black boxes". This is why Explainable AI (XAI) is becoming increasingly popular [28]. In addition to interpreting predictions and explaining the decision-making process, XAI is also used for model optimization to make machine learning models more computationally efficient. Such an approach is particularly useful in onboard space exploration applications where computational resources are very scarce [25]. Below are presented selected XAI methods.

SHapley Additive exPlanations (SHAP) is an approach aiming to explain the model's predictions [21]. The method is based on classic *Shapley values* from game theory which aims to estimate the contribution of each feature to the outcome [30]. The principle defines an explainable function g that explains the original prediction function, in the following formula:

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i \tag{1}$$

$z' \in \{0,1\}^M$ is *simplified input features* as derived in [21], where 1 means the feature exists in the coalition and 0 denotes the absence of the feature. M is the maximum size of the simplified input features. ϕ_i is the Shapley value for the i^{th} feature. Shapley values represent the importance of features and are derived in the following manner. The model is trained with all possible feature subsets $S \subseteq F$, where F is the set of all features. In order to evaluate the importance of a feature, the model $f_{S\cup\{i\}}$ is trained first with and then without that particular feature. Input features are accordingly $x_{S\cup\{i\}}$ and x_S. The difference between these

predictions is computed for all possible subsets $S \subseteq F \backslash \{i\}$ and averaged to assign the importance to the feature:

$$\phi_i = \sum_{S \subseteq F \backslash \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \tag{2}$$

The importance of features is frequently used to improve the prediction accuracy as well as to optimize the model. Temenos et al. employed SHAP for land use and land cover (LULC) classification in remote sensing [33]. Their pipeline included a deep convolutional neural network (CNN) and a SHAP. First, they predicted multichannel images using the CNN and then they fed the classification results with the images to the SHAP explainer. This way they understood which spectral band combinations of images performed better in the classification problem. Finally, using the SHAP explainer, they were able to demonstrate that the qualitative visual patterns help to increase classification accuracy. Another successful usage of SHAP is presented by Gurram et al. [17]. They addressed a problem of feature selection for non-linear kernel-based support vector machines, which tends to be a NP-hard problem. By deriving feature subsets using Shapley values and building the final model they improved the classification performance for hyperspectral datasets.

In deep learning, visualizing feature maps can be helpful in interpreting neural networks. However, as the number of layers increases, it becomes increasingly difficult to understand which patterns have been learned. Therefore, approaches to make the output of each layer more interpretable have been developed. One of the algorithms is *guided backpropagation* [32]. This technique allows the visualization of pixels in the input image that maximize the activation of neurons in higher layers of the network. The method is a combination of backpropagation and the deconvolutional network ('deconvnet') [38]. The idea is to revert the data flow from the specific high layer of the network back to the input image. Additionally, guided backpropagation restricts the flow of negative gradients by setting them to zero. As Shapley values, this technique can also be used in optimization by tuning the hyperparameters of the network according to the patterns learned by the convolutional layers.

3. Methodology

3.1. Data exploration

The Curiosity rover recently passed 4000 sols, i.e. Martian days, of time spent on Mars. During this time, ChemCam has collected data from more than 4000 targets. The existing metadata does not contain any information on whether dust displacement occurred during the LIBS measurements.

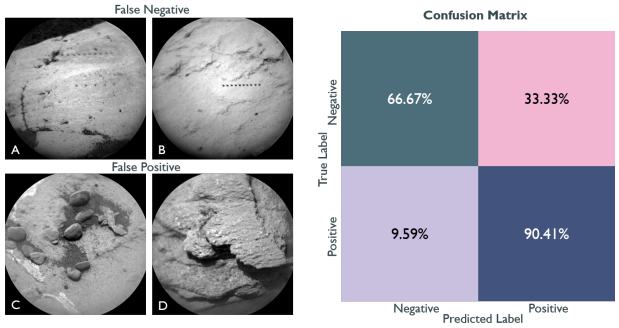

Figure 2. On the left erroneously predicted samples: A and B are false negatives and C and D False positives. On the right confusion matrix is normalized for each class.

Therefore, a visual inspection was carried out as a first step. By comparing the images taken before and after the LIBS measurements, and even by examining only the post-LIBS images, it was clear that the dust had shifted on some of the targets during the LIBS measurements as it is shown in Fig. 1. The initial challenge of the lack of unlabeled data was addressed by manual labeling. This allowed to employ a first simple binary classifier for further data exploration.

A total of 200 images taken after the LIBS measurements were manually labeled with dust displacement and the same amount without dust displacement, resulting in a total of 400 images. The images were retrieved from the PDS Geosciences node from sol 1000 to sol 3300 [36]. The size of each image is 1024×1024. Due to the small number of labeled samples, building a classifier from scratch would not be feasible. Therefore, a pretrained model was employed.

VGG16, trained on the ImageNet dataset, was selected as the pretrained model for our dataset [7, 31]. The model expects the input tensors to be of the size of $3 \times 299 \times 299$. So the post-LIBS images were read as RGB, by replicating each image across the RGB channels, and resized to meet the requirements. The labeled set was divided into training, test, and validation in the corresponding ratio of $70/15/15$. The training data was augmented by randomly flipping, rotating, and adjusting the brightness. No fine-tuning was performed at this point. The accuracy of the model reached 70%. After fine-tuning the model by training layers from the 11th layer onward with binary cross-entropy as the loss function, the accuracy increased to 82%. However, the learning curve showed overfitting, and the difference between training and validation set accuracy was about 10%, indicating a generalization problem. In addition, the learning curve showed fluctuations. These issues can be explained by the small size of the labeled dataset. Furthermore, false positive predictions occurred when there was an RMI image of a drill or a target with a rough texture, while false negatives occurred on RMIs where dust displacement

was clearly visible by visual inspection but the target surface was smooth. The left side of Fig. 2 shows examples of false negatives and false positives. Targets A and B have smooth texture and dust displacement is easily observed by the human eye. However, the model predicted that no dust displacement occurred. A few pebbles are visible in the target C and D has a nodular texture with layers. Both targets represent false positives. On the right side of Fig. 2, we show a confusion matrix of the VGG16 model after fine-tuning. The confusion matrix shows the ratio of correctly and incorrectly predicted samples, and indicates that false positives are more common. In addition to visual inspection, if the target is dusty, it is expected that it will be easier to move dust on a smooth surface than on a target with significant surface roughness, which may hinder dust displacement. Based on the evidence found from erroneously predicted samples, it was concluded that the learning process was influenced by the texture of the target. Considering the previous works showing the variability of ChemCam targets in texture, it was decided to remove the texture component from the training set to allow the model to focus on dust displacement [8, 9]. The following section presents the preprocessing steps including texture removal that are performed before the Martian in-situ data are fed into machine learning models.

3.2. Data preprocessing

Previously, it was observed that feeding raw images into the model leads to confusion influenced by the texture of the LIBS targets. Therefore, the goal of this section is to preprocess the input in such a way that the model focuses on dust displacement. This can be achieved by removing the texture information from the training set. The RMIs taken prior to the LIBS measurement can be useful in this case. The idea is to subtract the before and after LIBS images and reduce unnecessary information such as texture in the training data. The so-called difference images will show only the information about the change in the field of view.

Before proceeding to the difference images, it should be noted that there is an offset between the pre and post-LIBS RMIs, as shown in Fig. 1, due to the location of the RMI and LIBS instruments, which both reside in the rover's mast. As the LIBS measurements are made in a raster fashion, the head slightly changes its initial position. The offset can be described as a change along the x and y-axis. Although the movement of the mast introduces an angle, we did not include it in the preprocessing steps. In order to remove the offset, the RMIs before and after the LIBS are first cropped with respect to the offset coordinates. Subtraction is then possible, but due to the circular shape of the RMIs, some of the texture may be preserved in the corners. Therefore, after subtraction, the difference image is cropped to the center, keeping 70% of the image. The preprocessing procedure

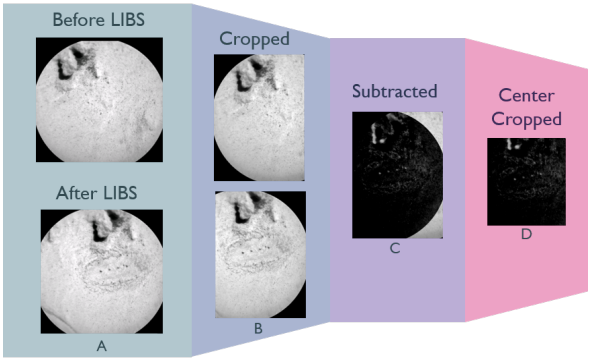

Figure 3. Schematic representation of preprocessing steps for RMI images. Target name Barouffieres. Section A represents the original input RMIs. In section B, the input images are cropped according to their offset values. Section C shows the difference image, which is centered in section D.

is illustrated in the Fig. 3, A being the input RMIs, B are inputs after being cropped according to the offset values, C showing the difference image, and D representing the center-cropped image. Finally, the dataset is ready to be divided into training, test, and validation sets after all these preprocessing steps. The data is split again in the ratio of $70/15/15$. The training set is augmented with several different filters such as random flip, random rotate, random hue, random saturation, brightness adjustment, contrast adjustment, and transpose.

3.3. Dust Displacement Detection Pipeline

The preprocessed dataset, i.e. the center-cropped difference images, was fed into the pretrained model of VGG16 repurposed for binary classifier by adding dense Softmax layers of two units and without any layers being trained on our data. The accuracy reached to 75%. The performance was better compared to the one without preprocessed data. Although the accuracy improved, it did not surpass the performance of the fine-tuned model on the data before preprocessing. Therefore, VGG16 was fine-tuned from layer 11 onwards on the preprocessed data and the accuracy increased to 84%. However, the learning curve kept displaying fluctuations and overfitting.

The problem of an overfitting model can be solved by increasing the labeled set or changing the model. Since manual labeling is tedious, it was decided to modify the classification model. We decided to use a model that can deal with overfitting such as a Random Forest [3]. Overfitting can be resolved if enough trees are used. We could feed the preprocessed training data directly into the Random Forest classifier, but this would have two drawbacks. First, flattening the images and feeding them directly into the classifier would increase the feature-sample ratio and cause the so called "*curse of dimensionality*" [2]. Second, as the previ-

ous work showed, the model would classify images according to their pixel values rather than the dust displacement pattern [9]. Therefore, we decided to keep part of the existing feature extraction model and add another classification model that works well with small data sets such as the described Random Forest. Specifically, transfer learning was still used for feature extraction, but the extracted features were no longer fed to the fully connected neural network, but to a binary Random Forest classifier. The downstream task of the Random Forest classifier was to detect whether or not dust displacement had occurred. Such an approach made it possible to take advantage of both a deep learning based pretrained feature extractor model and an ensemble learning model such as a Random Forest.

Three common pretrained models VGG16 [31], Inception V3 [10], and Resnet [18] were selected as feature extractors. Initially, no layers were trained, i.e. we froze all layers and kept the original weights. We attached a Random Forest classifier with a default number of 100 decision trees to the pretrained networks. The performance of all three pipelines was checked over multiple runs with different training and validation splits. All of them achieved higher than 86% accuracy. The performance is summarized in Tab. 1. As the models are trained on millions of images with 1000 different classes, they are all able to describe the training data well enough. Adding a Random Forest classifier improved model accuracy, even with default hyperparameters.

Pretrained Model	Attached Model	Accuracy
VGG16	RF	88.6%
Inception V3	RF	87.1%
Resnet	RF	86.6%

Table 1. *Performance of the pretrained VGG16, Inception V3, Resnet.*

Next we fine-tuned the pretrained models by training their last few layers on our training dataset. VGG16 showed better performance than other models even before fine-tuning. The reason why VGG16 is superior to other architectures can be explained by its size. As the training data is very small, it may be sufficient to extract its features using a smaller network rather than Inception V3 and Resnet, which go deeper and cause model confusion. This motivated us to explore VGG16 further.

During fine-tuning VGG16, we observed that freezing the first 10 layers outperformed other combinations. In order to explore those layers, we removed the rest of the layers and kept the original weights of those 10 layers. We also fine-tuned the attached random forest. We did a grid search starting from 100 trees to 2000. The fine-tuned Random Forest constructed from an ensemble of 1000 decision trees achieved an average accuracy of 92%, making it the best performing pipeline. Such a huge number of trees can be explained by the size of the input array. The input 1D array has a length of $16 \times 16 \times 256$, due to the flattened feature maps of size 16×16 and having 256 filters at the 10th layer of the VGG16. To compensate for the large number of input features the Random Forest classifier required 1000 trees.

Another classification model that can cope with a limited number of samples is Support Vector Machine (SVM) [15, 19]. We employed SVM with different kernels (linear, polynomial, sigmoid, and RBF) for the classification part of the extracted features [27]. Linear kernel SVM performed better than the others. In Tab. 2 we can observe that the combination of the whole VGG16 and a Random Forest outperforms the combination of VGG16 and a SVM by 18%. As mentioned above, we have 65536 features.

Pretrained Model	Attached Model	Accuracy
VGG16 10 layers	RF	92%
VGG16	RF	89%
VGG16 fine-tuned	RF	88%
VGG16 + PCA	RF	82%
VGG16 10 layers	SVM	80%
VGG16	SVM	71%
VGG16 10 layers + PCA	SVM	84%
VGG16 + PCA	SVM	83%

Table 2. *Performance of the pretrained VGG16 model in the combination of Random Forest (RF) and Support Vector Machine (SVM). PCA stands for principal component analysis.*

Feeding such a huge number of features can result in low model accuracy. Therefore, we reduced the input data dimensionality using Principal Component Analysis (PCA) [24]. The performance of the SVM improved, but it was still unable to achieve more than 84%. Therefore, we decided to keep Random Forest as our classification model. PCA was also applied on the input data for the Random Forest classifier. Cross-validation showed that the overall accuracy decreased. Random Forest already does the selection of feature subsets, and in addition, feature reduction may be unnecessary. The summary of the different models combined with VGG16 can be found in Tab. 2. All of the above steps helped us build the final pipeline, which consists of a preprocessing step according to Fig. 3, VGG16 with the first 10 pretrained layers for feature extraction, and a Random Forest classifier with 1000 trees. We call this pipeline Dust Displacement Detection (DDD) of which the schematic is shown in Fig. 4.

4. Experiments and results

The developed DDD pipeline was tested on the curated dataset and number of conclusions were drawn. On the left of Fig. 5, panel A is an example of a false negative predic-

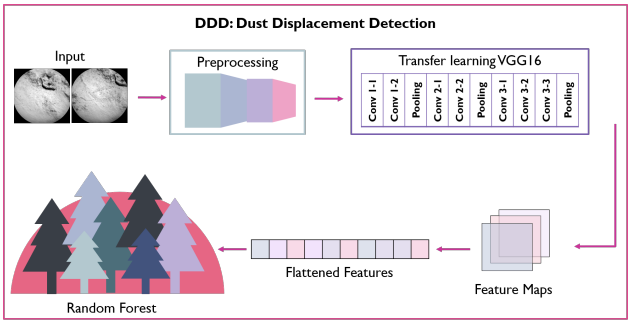

Figure 4. Schematic representation of Dust Displacement Detection (DDD) architecture.

tion that can occur when there is a large offset value and the area where the dust shift occurred is lost during center cropping. In panel B, an example of a false positive is illustrated. The false positive predictions can be caused by changes in the angle of the rover's mast. The preprocessing deals with an offset along the x and y axis, but does not take into account the change in angle in the field of view. The angle can introduce noise into the difference images, leading to model confusion. On the right is DDD's confusion matrix, showing how the model improved predictions over the one shown in Fig. 2. The high number of false positives observed initially is being resolved with the current pipeline.

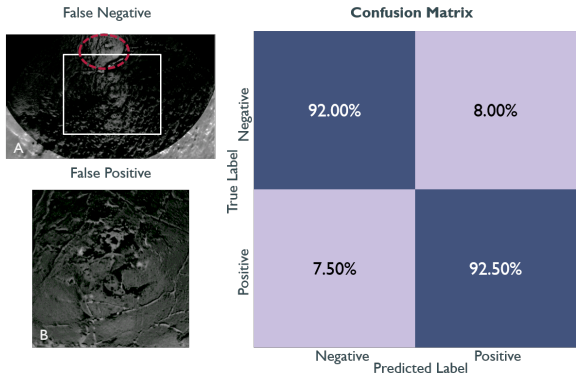

Figure 5. On the left, A represents a false negative sample before center cropping. The white contour showed the region after cropping and the dashed red contour encircles the dust displacement. B is a False Positive sample. On the right is shown the confusion matrix of DDD normalized for each class.

4.1. Importance of feature maps

Considering the limited memory and computing power of remote sensing platforms, we were motivated to check whether all the feature maps were necessary for the classification model and if there was room for model optimization by removing some of the feature maps. We evaluated whether some of the feature maps contribute more to the model accuracy and whether there is a subset of feature maps our model can afford to lose. Therefore, we estimated the importance of feature maps and ranked them by this parameter. In order to determine the minimum number of feature maps that is sufficient to achieve a model accuracy greater than 90%, we fed subsets of sorted feature maps and monitored the model performance.

The values of the feature maps varied according to the sample from which they were retrieved. In order to estimate the importance of feature maps that best described the data, we examined only the correctly predicted samples. The most activated feature maps, i.e. having the brightest pixels, were assumed to play a decisive role in describing the dataset. Following this reasoning, the feature maps for each sample were sorted by magnitude. The process followed the Algorithm 1. In the first step, for each sample, 16×16 pixel values of feature maps were summed up and sorted by magnitude. In the summing step, no preprocessing of the pixel values was done. All the values were positive allowing us to sum up them directly. Additionally, we did not normalize data as we wanted to keep the raw data to compare the feature maps. This allowed ordering feature maps by being most to least activated. Later, indices of the ordered feature maps were retrieved and stacked in a matrix M. In the second step, the matrix was examined. The vectors of ordered feature maps' indices differed slightly. We have selected the most common order of indices from matrix M for further investigation.

To verify the procedure, we explored the importance of features in Random Forest classifier and how they relate to the sorted feature maps derived from Algorithm 1. The importance of features was obtained from the explanation of the model's predictions using SHAP [21]. The idea was to check whether the features with the highest Shapley values, i.e. features that contributed most to the model outcome, belonged to the important feature maps derived using the Algorithm 1. In total, there are $16 \times 16 \times 256$ features. In order to map the features with the highest Shapley value to our ordered feature maps, we had to reshape flattened features back to 16×16. The results show that indeed the features with the highest importance belong to the highest magnitude feature maps. Fig. 6 shows Shapley values of pixels in feature maps sorted via the Algorithm 1. The violin plots of the Shapley values tend to decrease in magnitude as the importance of the feature maps decreases. This means that the feature maps extracted from VGG16 match the important features extracted from the random forest, i.e. the importance of features propagated throughout the whole architecture. This conclusion motivated further exploration of the feature maps.

To determine the number of the important feature maps that would be sufficient for dust displacement prediction

Algorithm 1 Sort Feature Maps by Importance

f represents pixel values of feature maps
T_c represents a set of correctly predicted test samples
M matrix of sorted indices of feature maps

Step 1:
▷ Sum and sort feature maps:
Function *SortFeatureMaps(f,j)*:
$N = 256$ number of filters
j index of a sample
for $i \in N$ **do**
allFeatureMaps.append$[\sum_{l,w=0}^{16} f_{l,w,i}]$
end for
$sortedIndices =$ **Sort**$(allFeatureMaps)$
Return $sortedIndices$

for $j \in T_c$ **do**
$M = Stack(SortFeatureMaps(f, j))$
end for

Step 2:
▷ find most common order of feature maps:
Function *IndicesSortedByImportance(M)*:
Return retrieve the most common vector from M

$IndicesSortedByImportance(M)$

we consider subsets of size 10 of the most activated feature maps as ranked in M and feed them sequentially to the Random Forest classifier. We check the accuracy of the Random Forest classifier by first feeding the classifier with the 10 highest magnitude feature maps and zeroing the rest. Then adding values of another 10 feature maps according to the ranked order and again observing the classification accuracy. We continue till all the feature maps have been considered. The results, Fig. 6, showed that only 85 out of the total 256 feature maps account for 90% of the model's accuracy. This means that the feature extraction part can be reduced by using only the required number of filters. This results in less computation for feature extraction and fewer feature maps to feed into the Random Forest. Minimizing the workload is especially important for remote sensing applications and our approach supports the discussion of machine learning models onboard.

To verify that these 85 feature maps are sufficient for the important feature extraction, we visualized what patterns have been learned by the output layer of the pretrained model. As the depth of the neural networks increases, the feature maps in the deeper layers become difficult to interpret. This is why we used *guided backpropagation* to illustrate the patterns learned by the optimized layer 10 of VGG16 [32]. This method visualizes patterns learned collectively by all feature maps in the layer. To visualize what was learned from just those 85 feature maps, the pretrained model had to be modified. The idea was to keep the 85 most important feature maps and set the weights for the rest to zero. Since layer 10 of VGG16 is a max pooling layer, we went one layer back to layer 9 and modified its weights. Afterward, we attached the max pooling layer and were back to the optimized model. In Fig. 7, two examples of the original input images are provided in panel A and the visualization of the learned pattern by the optimized 10th layer in panel B. The brightest pixels translate into the most activated ones. In panel B, one can detect pixels activated in the area of dust displacement. Although there are white dots as learned noise, the dust contour in the center is still noticeable. The noise may be caused by the offset angle in difference images and the texture of the target, and possibly changing light conditions between the pre- and post-LIBS images may be the contributor to the learned patterns. Over-

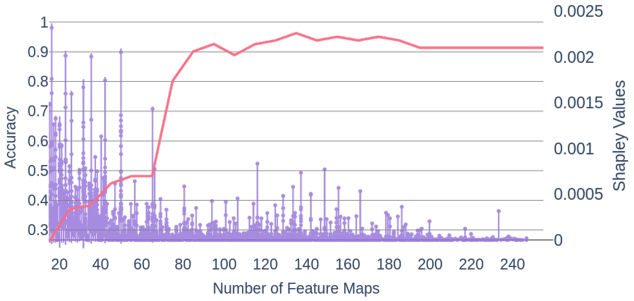

Figure 6. Red line displays the dependency of the model accuracy to the number of feature maps. Purple violin plots show Shapely values of feature maps sorted according to importance.

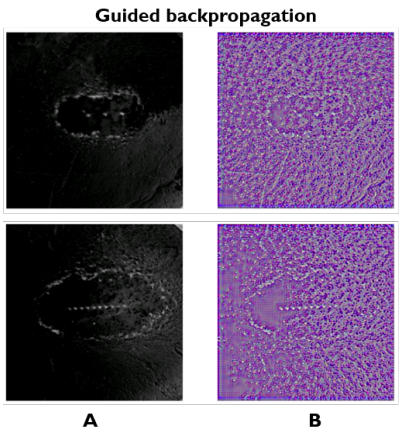

Figure 7. Section A represents two different preprocessed samples while images in section B are the same samples after applying the guided backprogation.

all, the interpretation of patterns using guided backpropagation helps to validate the statement that reducing the number of feature maps to the most important ones is feasible to detect dust displacement.

5. Applications

Since manual interpretation of data is tedious and does not prevent human error, the use of machine learning methods comes in handy, allowing experts to interpret results efficiently rather than going through all the targets. Two applications of our pipeline for Mars data are further discussed. Fig. 8 showcases 1925 targets from the initial sols up to sol 3778 and whether dust displacement occurred while taking LIBS measurements. A darker shade indicates that no dust displacement is occurring in the corresponding sol. The plot shows that dust displacement is observed less frequently in the early sols when the rover was on the crater floor and foothills of Mt. Sharp than later, higher on Mt. Sharp.

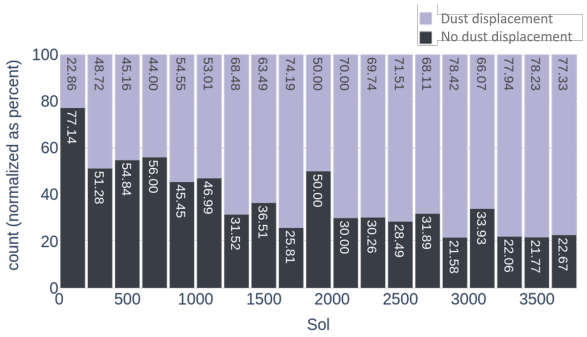

Figure 8. Bar plot showcasing percentile ratio of dust displacement sol wise. Each bin contains 200 sols. In total 1925 targets are processed.

Another use case of the model is to show whether there is a dependence of dust displacement on the distance between a target and the laser. Fig. 9 shows that less dust displacement occurs with an increase of the distance. This could also point to a change in laser-target coupling so that less energy is transferred into the LIBS plasma and it's expansion, i.e. the shockwave. This was already observed for the LIBS data as the LIBS signal intensity decreases with the increase of the distance [1]. Another factor in weak coupling could be dust, which is not removed by a shock wave because the laser is further away from the sample. However, it should be noted that the number of measurements with a higher distance is very low. There are only 4 targets measured from the distance of $6 - 6.5$ meters. Although these plots alone do not allow us to draw distinct conclusions, they do encourage further research. The study can be continued with dust contour detection and the relation of dust movement shape to the distance of the laser or texture of the target.

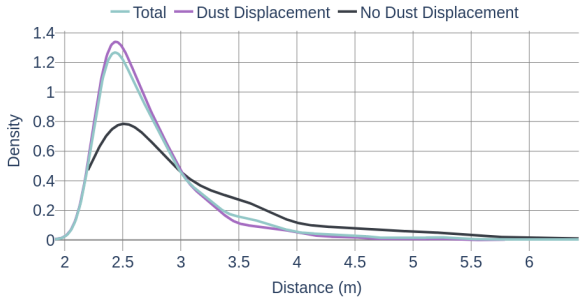

Figure 9. Dust displacement in relation to the distance of the laser to the target. Distance is measured in meters. Area below the curves are normalized to 1.

Additionally, the dust detection pipeline has the potential to be modified for the data acquired by the SuperCam instrument on the Perseverance rover. The instrument is also made up of LIBS and RMI components and is measuring targets in Jezero crater for three years now [23]. The SuperCam RMI captures RGB images, which could potentially enhance the DDD.

6. Conclusion

We propose a pipeline to automatically detect dust displacement on ChemCam targets using data acquired by the RMI before and after LIBS measurements. We preprocessed the RMIs and explored the data to understand the classification model requirements. Due to the limited size of a labeled dataset for training, we chose to employ transfer learning. We tested different pretrained models and combinations of Random Forest and SVM to find the best performing two-stage pipeline. The final pipeline consisted of the first 10 layers of VGG16 for feature extraction and a Random Forest binary classifier. The accuracy of the model reached 92%. For optimization purposes, we estimated the importance of the feature maps, verified their contribution using SHAP and guided backpropagation, and kept only those that had a significant impact on the accuracy of the pipeline. Out of a total of 256, only 85 feature maps were found to contribute to 90% classification accuracy, thus indicating a possible reduction of computation for onboard processing of martian in-situ data. We thus addressed the classic challenges of planetary exploration, namely limited labeled data and scarce computational resources. By showing application examples, a few points were concluded: DDD allows researchers to efficiently detect the dust displacement on targets analyzed along the entire traverse, as well as to explore the phenomena for each target and derive its dependence on target or instrument parameters. Overall, the work supports the use of machine learning based methods for current Martian studies and encourages further research.

References

[1] R. C. Wiens, S. Maurice, B. Barraclough, M. Saccoccio, and W. C. Barkley. The ChemCam Instrument Suite on the Mars Science Laboratory (MSL) Rover: Body Unit and Combined System Tests. *Space Science Reviews*, 170:167–227, 2012. 8

[2] N. Altman and M. Krzywinski. The curse(s) of dimensionality. *Nature Methods*, 15:397–400, 2018. 4

[3] L. Breiman . Random Forests. *Machine Learning*, 45:5–32, 2001. 4

[4] B. Chide, S. Maurice, N. Murdoch, J. Lasue, B. Bousquet, X. Jacob, A. Cousin, O. Forni, O. Gasnault, P. Y. Meslin, J. F. Fronton, M. Bassas-Portús, A. Cadu, A. Sournac, D. Mimoun, and R. C. Wiens. Listening to laser sparks: a link between laser-induced breakdown spectroscopy, acoustic measurements and crater morphology. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 153:50–60, 2019. 2

[5] S. M. Clegg, R. C. Wiens, R. Anderson, O. Forni, J. Frydenvang, J. Lasue, A. Cousin, and V. Payré et al. Recalibration of the mars science laboratory chemcam instrument with an expanded geochemical database. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 129:64–85, 2017. 2

[6] D. A. Cremers and L. J. Radziemski. *Handbook of Laser-Induced Breakdown Spectroscopy*. John Wiley Sons Ltd, third edition, 2006. 1

[7] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 2, 3

[8] A. Essunfeld *et al*. Attribute recognition for grouping elevated-manganese ChemCam targets by visual characteristics. *52nd LPCS*, 2021. 2, 4

[9] A. Lomashvili *et al*. Rock classification via transfer learning in the scope of ChemCam RMI image data. *LPSC*, 2023. 2, 4, 5

[10] C. Szegedy *et al*. Rethinking the Inception Architecture for Computer Vision. *IEEE CCVPR*, 2016. 2, 5

[11] J. Lasue *et al*. Martian Eolian Dust Probed by ChemCam. *Geophysical Research Letters*, 2018. 1, 2

[12] R. C. Wiens *et al*. The ChemCam Instrument Suite on the Mars Science Laboratory (MSL) Rover: Body Unit and Combined System Tests. *SSR*, 170, 2012. 1

[13] S. Maurice *et al*. The ChemCam Instrument Suite on the Mars Science Laboratory (MSL) Rover: Science Objectives and Mast Unit Description. *SSR*, 170, 2012.

[14] S. Mouélic *et al*. The ChemCam Remote Micro-Imager at Gale crater: Review of the first year of operations on Mars. *Icarus*, 249, 2015. 1

[15] M. Fernandez-Delgado, E. Cernadas, and S. Barro. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research*, 15:3133–3181, 2014. 5

[16] J. P Grotzinger, J. Crisp, A. R. Vasavada, and R. P Anderson. Mars Science Laboratory Mission and Science Investigation. *Space Science Reviews*, 170:5–56, 2012. 1

[17] P. Gurram, H. Kwon, and C. Davidson. Shapely Value Based Random Subspace Selection for Hyperspectral Image Classification. *IGARSS*, 2015. 3

[18] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 5

[19] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, and B. Scholkopf. Support vector machines. *IEEE Intelligent Systems and their Applications*, 13(4):18–28, 1998. 5

[20] V. Lakshmanan, M. Görner, and R. Gillard. *Practical ML for Computer Vision*. O'Reilly, third edition, 2021. 2

[21] S. Lundberg and S. I. Lee. A Unified Approach to Interpreting Model Predictions. *NIPS 2017*, 2017. 2, 6

[22] S. Maurice, S. M. Clegg, R. C. Wiens, and O. Gasnault. ChemCam activities and discoveries during the nominal mission of the Mars Science Laboratory in Gale crater, Mars . *J. Anal. At. Spectrom*, 31:863–889, 2016. 1

[23] S. Maurice, R. C. Wiens, and P. Bernardi et al. The SuperCam Instrument Suite on the Mars 2020 Rover: Science Objectives and Mast-Unit Description. *Space Science Reviews*, 217, 2021. 8

[24] A. Maćkiewicz and W. Ratajczak. Principal components analysis (pca). *Computers Geosciences*, 19(3):303–342, 1993. 5

[25] D. Minh, H. X. Wang, Y. F. Li, and T. N. Nguyen. Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, 55:3503–3568, 2022. 2

[26] C. A. Moore, J. E. Moores, C. E. Newman, M. T. Lemmon, S. D. Guzewich, and M. Battalio. Vertical and horizontal heterogeneity of atmospheric dust loading in northern gale crater, mars. *Icarus*, 329:197–206, 2019. 2

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 5

[28] R. Roscher, B. Bohn, M. F. Duarte, and J. Garcke. Explainable machine learning for scientific insights and discoveries. *IEEE Access*, 8:42200–42216, 2020. 2

[29] F. Seel, S. Schröder, D. S. Vogt, E. Dietz, H. W. Hübers, and M. Gensch. Generation and evolution of laser-induced shock waves under Martian atmospheric conditions . *Icarus*, 394, 2023. 2

[30] L. S. Shapley. A value for n-person games. *Contributions to the Theory of Games*, 2.28:307–317, 1953. 2

[31] K. Simonyan and A. J. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ICLR 2015*, 2015. 3, 5

[32] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller. Striving for Simplicity: The All Convolutional Net. *ICLR 2015*, 2015. 3, 7

[33] A. Temenos, N. Temenos, M. Kaselimi, A. Doulamis, and N Doulamis. Interpretable Deep Learning Framework for Land Use and Land Cover Classification in Remote Sensing Using SHAP. *IEEE Geoscience and Remote Sensing Letters*, 20, 2023. 3

[34] A. R. Vasavada. Mission Overview and Scientific Contributions from the Mars Science Laboratory Curiosity Rover After Eight Years of Surface Operations. *Space Sci Rev*, 218, 2022. 1

[35] A. Vicente-Retortillo, G. M. Martínez, and N Renno. Seasonal Deposition and Lifting of Dust on Mars as Observed by the Curiosity Rover. *Sci Rep*, 8, 2018. 1

[36] R. C. Wiens. Msl chemcam remote micro- imaging camera raw data. MSL-M-CHEMCAM-RMI-2-EDR-V1.0, NASA Planetary Data System, 2013. 3

[37] A. Witze. Mars rover finds evidence of ancient habitability. *Nature*, 2013. 1

[38] M. D. Zeiler and R. Fergus. Visualizing and Understanding Convolutional Networks. *ECCV*, 2014. 3